Н новости фундаментальной физики

фф

выпуск 2

КВАНТОВАЯ ГРАВИТАЦИЯ И ТОПОЛОГИЯ

ИЗДАТЕЛЬСТВО «МИР»

Москва • 1973

новости фундаментальной физики

выпуск

2

КВАНТОВАЯ ГРАВИТАЦИЯ И ТОПОЛОГИЯ

СБОРНИК СТАТЕЙ

Перевод с английского
канд. физ.-мат. наук Б. Н. Фролова

Под редакцией
проф. Д. ИВАНЕНКО

ИЗДАТЕЛЬСТВО
«МИР»

МОСКВА • 1973

УДК 531.51 + 523.11

Сборник содержит статьи видных зарубежных ученых, посвященные ряду актуальных проблем современной теории гравитационного поля и его связи с квантовой физикой элементарных частиц. Главное внимание уделяется принципиальным вопросам структуры пространства-времени, свойствам решений эйнштейновских уравнений общей теории относительности, обзору методов квантования гравитации. Вместе с тем в сборнике затрагиваются более конкретные вопросы коллапса и структуры частиц. В итоге начинают вырисовываться контуры объединенной картины пространства-времени, гравитации и микромира элементарных частиц.

Сборник рассчитан на широкий круг физиков, математиков, астрономов, философов, интересующихся вопросами современной гравитации, астрофизики, космологии.

Редакция литературы по физике

К $\frac{0232\text{-}067}{041(01)\text{-}73}$

Вступительная статья

Гравитация и единая теория

Д. ИВАНЕНКО

§ 1. ВВЕДЕНИЕ

1. Гравитационные конференции

Новейшие проблемы релятивистской гравитации, в особенности связанные с космологией, астрофизикой, Солнечной системой, элементарными частицами и гравитационными волнами, вышли на первый план в физике и астрономии и начинают все больше интересовать ученых в более далеких от гравидинамики областях — исследователей космоса, геологов и других. Значительно увеличилось количество конференций, симпозиумов, школ.

Представительные конференции собираются каждые три года по инициативе Международного гравитационного комитета. После 5-й конференции 1968 г. в Тбилиси в центре внимания 6-й конференции (Копенгаген, 1971 г.) стояли доклады по топологии, квантовой гравитации, уточненному исследованию Солнечной системы, поискам гравитационных волн, проблеме коллапса и другим.

Каждые два года в США созываются симпозиумы по релятивистской астрофизике; шестой подобный «Техасский» симпозиум состоялся в декабре 1972 г. в Нью-Йорке. Со своей стороны Советская гравитационная комиссия продолжает организацию симпозиумов, школ и больших гравитационных конференций, первая из которых состоялась в 1961 г. в Москве, вторая — в 1965 г. в Тбилиси, третья — в октябре 1972 г. в Ереване.

В 1972 г. имели место следующие встречи гравитационистов. В Сицилии в мае проходили занятия высшей гравитационной школы, посвященные космологии. Римская Академия наук в июне организовала школу по релятивистской астрофизике. В Кембридже летом состоялся симпозиум, посвященный экспериментам в общей теории относительности (ОТО). Один из трех циклов летней высшей школы имени Ферми в Варенне был посвящен ОТО. В апреле в Вартбурге (ГДР) состоялся общерелятивистский симпозиум. Летние гравитационные школы были организованы в Лезуш (Франция) и в Симле (Индия).

Отметим еще гравитационную конференцию 1969 г. в Цинциннати, из трудов которой в настоящий сборник включен доклад Герока. Кроме того, следует иметь в виду гравитационные доклады на конференциях по элементарным частицам. Позднейший вариант доклада Салама на 15-й («Рочестерской») конференции по физике высоких энергий в Киеве, представленный им на сессии Американского физического общества, мы публикуем в нашем сборнике.

Пятисотлетие со дня рождения Коперника будет отмечено в Польше гравитационными конференциями в сентябре 1973 г.; этой замечательной дате посвящается симпозиум в Киеве в феврале 1973 г.

Перед нами возникает внушительная картина растущего научного и общественного интереса к разнообразным проблемам гравитации. Этот интерес можно объяснить тремя причинами. Во-первых, он связан прежде всего с глубоким характером эйнштейновской ОТО, остающейся одним из надежных фундаментов современной картины мира в смысле понимания структуры пространства-времени, его связи с веществом, а также в смысле трактовки новейших астрономических объектов типа квазаров и пульсаров, в направлении уточнения небесной механики и, наконец, в смысле единственно правдоподобного наброска космологии, т. е. взрывного происхождения, эволюции и судьбы всей известной вселенной.

Во-вторых, как стало ясно в последние годы, новейшие астрономические открытия и возможность релятивизации небесной механики Солнечной системы и уточненные лабораторные эксперименты перевели ОТО из принципиальной многозначительной системы, покоившейся всего на трех эффектах и фридман-хаббловском расширении вселенной, в разряд более реальных областей физики и астрономии. В-третьих, неуклонно, хотя и медленно, происходит смыкание квантовой теории элементарных частиц с гравитацией и космологией.

История гравитации, вновь привлекающая усиленное внимание, обсуждалась на 11-м и 12-м конгрессах по истории науки (Париж, 1968 г., Москва, 1971 г.). Отсылая за деталями к содержательному докладу Гута [1] по истории ОТО на конференции в Цинциннати, к монографии Тоннела [2] об истории принципа относительности, к поучительной и волнующей переписке Эйнштейна и Зоммерфельда, изданной Херманном, и к тезисам докладов Хольтона, Тредера, Кудрявцева, Спасского и др. на упомянутых конгрессах, остановимся на двух недавних книгах Тредера [3, 4]. Анализируя принцип эквивалентности, автор подробно останавливается на полемике Эйнштейна с Абрагамом, Нордстремом и другими в годы создания ОТО, которая перекликается как с позднейшими выступлениями Фока о смысле принципа относительности, так и с конформно-плоскими вариантами каталога Торна (§ 2). Тредер развивает идеи механики Римана — Вебера, в которой масса тела зависит от взаимодействия с окружающими телами, строит модель типа ОТО, в которой, однако, механическое влияние тел на инерцию учитывается уже в ньютоновском приближении. На наш взгляд, здравое зерно в подобных махианских идеях следует искать прежде всего в возможном влиянии космологических асимметрий на «вакуум» (основное состояние) элементарных частиц: отсюда в духе теоремы Гольдстона возникают частицы — фотон, гравитон и другие. В частности, гравитон, трактуемый как «гольдстон» (Филипс, Гейзенберг, Иваненко, Тредер, Дюрр), соответствует искривлению плоского 4-пространства, т. е. нарушению лоренцевой симметрии.

2. Задачи сборника

Дадим краткие пояснения к статьям сборника, который открывается обзором известного молодого тополога Герока, посвященным важной проблеме сингулярностей в ОТО. Обсуждаются сингулярные решения ОТО и важное место, которое они занимают среди всех решений. Неизбежность сингулярностей перекликается с теоремами Хокинга — Пенроуза о неизбежности коллапса при весьма общих условиях. Подробно анализируется понятие «геодезической неполноты». В смысле Герока G-сингулярны модель Фридмана, решения Шварцшильда и Керра, тогда как пространство Минковского и плосковолновые решения не сингулярны (грубо говоря, в них геодезические продолжаются неограниченно). Положительным пунктом статьи является учет космологического члена, существенно влияющего на сингулярности. Мы не устаем повторять о необходимости его учета как из общих соображений, так и на основе эмпирических сведений; как видим, топология подтверждает эту точку зрения вопреки ряду авторов (Ландау — Лифшиц, Пахнер и др.), неоправданно отбрасывающих этот член.

В этой связи отметим доклады топологического симпозиума в Гватте — Берне 1970 г.: Лихнерович [5] анализировал «расслоение пространства» в связи со спинорными структурами. Зейферт [6] рассмотрел «определение каузальной границы», развивая идеи Пенроуза [14]. Упорядочение во времени рассмотрел Кронгеймер [7]. К статье Герока примыкает доклад Шмидта [8] «Новое определение сингулярных точек в ОТО». Пахнер [9], рассматривая конкретные решения, подчеркивает неизбежность сингулярностей в случае некогерентного вещества даже при наличии вращения.

В обширном докладе Картера [10] анализируется каузальная структура пространства-времени; в докладе одного из видных топологов Хокинга [11] рассматривается «метапространство» всех лоренцевых метрик на 4-многообразии. Эти доклады опубликованы в новом журнале «Общая теория относительности и гравитация» (GRG), издающемся с 1971 г. на английском языке в Берне под руководством Международного гравитационного комитета.

Резюме докладов конференции 1971 г. в Копенгагене дает Камезинд [18]. В докладе Герока вновь рассматривалась причинность (отсутствие замкнутых времениподобных кривых), стабильная каузальность пространства-времени, связь ориентируемости с *CPT*-теоремой. Важность топологических соображений, необходимых для трактовки космологии, проблем коллапса, квантовых флуктуаций метрики, теории суперпространства и других, ныне не подлежит сомнению.

Вторая статья сборника — большой обзор Брилла и Гоуди, посвященный трудным проблемам квантования ОТО, в которой квантованная метрика тесно связана с самим пространством-временем. Здесь снова ясно видно, как вопросы топологии, квантования и более конкретных аспектов космологии теснейшим образом переплетаются друг с другом. Один из параграфов обзора посвящен теории Мизнера [12], который рассматривает квантовую модель ОТО и ряд космологических выводов специально в применении к ситуации вблизи «большого взрыва». В установленных им уравнениях типа Гамильтона — Якоби и Шредингера параметры космологических моделей любопытным образом играют роль «потенциалов». (Характерные пульсации пространства-времени близ сингулярной точки анализировались также советскими авторами [19].) Статья Брилла и Гоуди в основном посвящена принципиальным проблемам и оставляет в стороне вопросы предсказанных нами трансмутаций, порождения и рассеяния гравитонов как квантов слабого поля во множестве процессов.

Наконец, в последней, третьей статье сборника Салам проводит учет взаимодействия гравитации с элементарными частицами, утверждая, что таким путем удается устранить столь известные расходимости в теории поля, унаследованные еще от классической теории точечного электрона Лоренца и сохранившиеся в квантовой электродинамике. Грубо говоря, основная идея заключается в том, что электростатические силы расталкивают, а гравитация сжимает «части» электрона. Существенным у Салама является использование earlier неполиномиальных лагранжианов, которые эффективно вводят члены, описывающие затухание. Хотя многочисленные работы Салама еще не получили безоговорочного признания, но, кроме ряда математических результатов, они несомненно содержат стимулирующие идеи, в том числе в направлении учета «сильной» гравитации, реализуемой мезонами спина 2.

Ниже мы даем обзор ряда принципиальных вопросов, в частности правдоподобных обобщений ОТО и информацию о новых экспериментальных данных в этой области; при этом астрофизические и космологические вопросы, как правило, не затрагиваются.

§ 2. НОВЫЕ ЭКСПЕРИМЕНТЫ В ОБЩЕЙ ТЕОРИИ ОТНОСИТЕЛЬНОСТИ

1. Релятивистская небесная механика

В согласии с предсказаниями ОТО красное смещение спектральных линий на Солнце было подтверждено с точностью до 5%. Точность 1% была достигнута в лабораторных условиях путем использования мессбауэровских гамма-фотонов (Паунд).

Отклонение в поле Солнца радиоволн, испущенных квазаром 3С279, который покрывается Солнцем каждый год 8 октября, измерялось с 1969 г. при помощи радиоинтерферометров (с базами от 1 до 4000 км), причем для сравнения служил квазар 3С273, расположенный на угловом расстоянии 10°. С довольно хорошей точностью были получены значения 0,90($\pm$0,05) [23] и 1,03($\pm$0,2) [20] (в долях эйнштейновского угла $4GM/l = 1{,}75R/l$). Эти прекрасные результаты дают возможность систематического наблюдения отклонения ОТО независимо от капризов погоды при затмениях, нарушавших и без того менее точные наблюдения в оптической области.

Замедление времени в ОТО. Наблюдение радиоэха, отраженного от Венеры или Меркурия, находящихся в верхнем соединении, когда радиоволны проходят вблизи солнечного диска, позволило измерить замедление времени, вызванное гравитационным полем Солнца. Согласно ОТО,

$$\Delta t = \left(\frac{4r_0}{c}\right) \ln \frac{r_l + r_p + R}{r_l + r_p - R},$$

где $r_0 = 1{,}5$ км, r_l, r_p и R обозначают соответственно расстояния Земля — планета, планета — Солнце и Земля — Солнце. Группа Шапиро [20—22] получила пассивным методом отражения сигнала значение $1{,}025 \pm 0{,}05$; Мулеман и др. [24], используя активную ретрансляцию при помощи спутников Маринер VI и VII, проходивших в верхнем соединении весной 1970 г., пришли к значению $1{,}00 \pm 0{,}04$.

Сдвиг перигелия. Анализ астрономических данных [20—22] для Меркурия привел к значению $\lambda = 1{,}00 \pm 0{,}01$. Одновременно анализировалась возможная сплющенность Солнца, которая оказалась равной $J_2 = (0{,}8 \pm 0{,}8) \cdot 10^{-5}$ [предсказание Дикке давало $J_2 = (2{,}7 \pm 0{,}5) \cdot 10^{-5}$].

Важно отметить, что при анализе результатов экспериментов, значений сдвига перигелия Меркурия и постоянства гравитационной константы группа Шапиро подвергла впечатляющей проверке разнообразнейшие параметры Солнечной системы (расстояния, массы, радиусы планет, сдвиги перигелия и др.), используя мощные компьютеры для обработки примерно 400 000 (!) астрономических данных, накопленных за 220 лет наблюдений с 1750 по 1970 г.

Пришлось произвести сложную работу по сравнению различных звездных каталогов, по взаимному сравнению и редукции различных констант при определении нутации, прецессии и т. д. По-видимому, прежние оценки сдвига перигелия Меркурия были слишком оптимистическими, и разумные границы точности составляют всего 3% в общем согласии с ОТО.

Анализ возможного изменения гравитационной константы со временем не дал положительного результата, во всяком случае

$$\frac{\dot{G}}{G} \lesssim 4 \cdot 10^{-10}\text{год}^{-1}.$$

Ожидается увеличение точности измерений запаздывания до 0,3% благодаря применению усовершенствованной радарной системы; продолжение измерения запаздывания Земля — Меркурий в течение 5 лет позволит снизить возможное изменение гравитационной константы до $3 \cdot 10^{-11}$ год$^{-1}$. Использование спутников типа «Викинг», запуск которых в сторону Марса планируется в 1975 г., по-видимому, позволит измерять эффект запаздывания с точностью 0,03%, так что в конце концов в 80-х годах окажется возможным обнаруживать эффекты второго порядка в ОТО.

Перечислим другие эффекты, измерение которых приобрело ныне актуальность ввиду различия предсказаний ОТО, скалярно-тензорной теории (СТТ) и других вариантов гравидинамики:

1. Геодезическая прецессия (гироскопа на спутнике Земли по отношению к далеким звездам).

2. Эффект Лензе — Тирринга «увлечения» вращающейся Землей инерциальной системы отсчета (приводящий к особой прецессии гироскопа).

3. Нарушение принципа эквивалентности в СТТ (Йордана — Бранса — Дикке) и ряде других неэйнштейновских вариантов гравидинамики приводит к множеству предсказываемых эффектов; к ним относятся (Нордтведт) «поляризация» орбиты Земля — Луна, поскольку ожидается большее значение ускорения к Солнцу Луны, чем Земли. Уточненное измерение расстояния Земля — Луна с помощью советско-французских и американских лазеров, помещенных на Луне, в принципе позволило бы открыть (или закрыть) этот эффект.

4. Различные периодичности в гравиметрических отсчетах (сверх основных 12-часовых периодов, обусловленных приливным влиянием Солнца и Луны и учитываемых уже в ньютоновской теории).

5. Возможные изменения гравитационной константы не только со временем благодаря расширению вселенной, но также благодаря влиянию других масс (последнее в принципе измеряется гравиметром и экспериментами типа Кэвендиша).

6. Различая более тонко понятия масс [инертной, пассивной, активной (Бонди) и «сохраняющейся» (Торн), которая в качестве массы покоя фигурирует в законах сохранения ОТО], следует задаться вопросом о значении отношений m_A/m_I и т. д. Опыты Этвеша — Дикке подтвердили постоянство m_A/m_I с точностью $3\cdot 10^{-11}$ независимо от химического состава тел. Различные обобщения ОТО предсказывают возможную зависимость m_A/m_C от химического состава, анизотропию активной массы m_A и т. д. Однако эксперименты Крейцера 1968 г. подтвердили постоянство m_A/m_P с точностью до $5\cdot 10^{-5}$ для брома и фтора. В принципе следует проверить роль электромагнитной энергии, ядерной энергии связи и других видов внутренней энергии в качестве источников гравитационного поля.

2. Гравитационные волны

Программа Вебера поисков гравитационных волн из космоса, доложенная им на конференции 1959 г. и описанная в его известной книге, сначала не привлекла особого внимания. Однако публикация первых же положительных результатов в 1969 г. вызвала буквально взрыв интереса. Отсылая за всеми деталями к обширной литературе, обрисуем кратко ситуацию к началу 1973 г. Детектором Вебера является алюминиевый цилиндр весом 1,5 т, длиной 1,5 м, тщательно изолированный от внешних воздействий (сейсмических, электромагнитных и других помех). Незначительные сдвиги цилиндра (10^{-14} см) регистрируются при помощи пьезоэлектрических или сегнетоэлектрических пластин, прикрепленных к нему. Принимаются в расчет всплески определенной значительной интенсивности над уровнем шумов, одновременно регистрируемые вторым таким же цилиндром, расположенным на расстоянии около 1000 км (частоты 1661 и 1581 Гц).

Несмотря на множество проектов других гравитационных детекторов, ряд лабораторий США, Англии, Советского Союза, а также в Риме — Фраскати и Мюнхене предпочел произвести прямую проверку опытов Вебера, смутивших всех необычайно высоким значением интенсивности предполагаемого гравитационного излучения, которое соответствует аннигиляции примерно 1000 солнечных масс в год, если его относить к центру Галактики (согласно указаниям на обнаруживаемую анизотропию). Было сделано немало критических замечаний по установке и обработке данных Вебера, который, однако, продолжал настаивать на реальности своих результатов [26—28].

В дальнейшем Вебер предполагает измерить квадрупольные колебания Луны с помощью специального гравиметра, который был послан с экспедицией «Аполлон-17» в декабре 1972 г. Фэйрбенк надеется, используя сверхпроводящую технику, изучить сдвиги торцов цилиндра Вебера с точностью до 10^{-14} см. В ряде лабораторий стали искать корреляции веберовских всплесков с колебаниями солнечной активности, сейсмикой и т. д.

Обратимся теперь к теоретическим аспектам проблемы гравитационных волн. Здесь следует различать, с одной стороны, вопрос об отыскании точных и приближенных решений ОТО волнового типа, согласно тем или иным критериям (Пирани, Бэл, Лихнерович, Зельманов — Захаров, Молдыбаева — Николаенко и др.), и, с другой стороны, проблему энергии, переносимой гравитационными волнами. Большинство кинематических критериев согласуется с тем, что гравитационные волны находятся среди точных решений вырожденного класса II классификации Петрова, хотя недавно Силенко [62] интерпретировал в волновом смысле также решение класса III. Хорошую сводку этой стороны дела дал недавно Захаров [29]. Как известно, несмотря на ряд аналогий с электромагнитными волнами, здесь имеются существенные отличия, обусловленные тензорным характером поля (спин 2; в низшем приближении квадрупольного, а не дипольного). Несмотря на трудный и далеко не решенный вопрос об энергии гравитационного поля, подавляющее большинство авторов согласно с тем, что гравитационные волны переносят энергию, которая определяется в низшем приближении квадратом третьей производной квадрупольного момента. Вопрос о гравитационном затухании рассматривал Торн.

В недавнем обзоре Торн и Пресс вновь анализируют возможные источники гравитационных волн. Если оставить в стороне ядерные взрывы и проект квантовых установок Копвилема, то имеются в виду только астрофизические и космологические источники: системы двойных звезд, пульсары, взрывы сверхновых и образование нейтронных звезд, взрывы в квазарах и ядрах галактик, конденсация галактик, различные случаи коллапса звезд, излучение гравитационных волн при предсказываемой аккреции вещества гигантскими черными дырами в ядрах галактик; наряду с этими процессами, рассмотренными Мироновским, Зельдовичем — Новиковым, Уилером — Руффини, Мизнером, Торном и др., имеется класс процессов, связанных с трансмутациями фотонов в гравитоны и порождением гравитонов в различных квантовых процессах (Иваненко, Соколов, Кармели, Хальперн, Герценштейн и др.). В начале 1973 г. наилучшим кандидатом в черные дыры (не открытые еще с полной достоверностью) считалась невидимая компонента объекта в созвездии Лебедя (Cygnus X-1), по-видимому представляющего собой двойную звезду, в которой перетекание вещества с видимой компоненты дает наблюдаемые вспышки рентгеновского излучения.

Наряду с этим в пространстве могут иметься гравитационные волны космологического происхождения, обусловленные процессом первичного «большого взрыва». Необычайно большая интенсивность гравитационных волн, подсказываемая результатами Вебера, привела к усиленным поискам необычных источников и разнообразных моделей галактического центра. Среди них упомянем предсказываемые черные дыры шварцшильдовского или керровского типа; Бертотти предполагает наличие в Галактике релятивистского скопления коллапсированных тел, которые «сжимают» свою гравитационную потенциальную энергию. Тем самым предлагается решить проблему недостающих масс главным образом за счет гравитационных волн и коллапсированных галактик. Кафка в качестве возможных источников веберовских сигналов предложил всплески гравитационных волн плотными скоплениями черных дыр («граварами») [30а, 31].

Наконец, отметим предложенный недавно механизм синхротронного излучения гравитационных волн. В самом деле, массивный объект, вращающийся с релятивистской скоростью вокруг черной дыры, должен испускать гравитационное излучение, в ряде отношений подобное электромагнитному синхротронному излучению, испускаемому релятивистскими электронами в магнитных полях ускорителей и астрономических объектов. Основные характеристики синхротронного излучения (резкая направленность, максимум спектра в области высоких гармоник, особая поляризация) хорошо изучены, и это позволяет предвидеть некоторые аналогичные свойства в гравитационном случае; в частности, прожекторный эффект устранил бы необходимость суммировать веберовские всплески по всем углам. Мы с Познаниным проанализировали гравитационное синхротронное излучение (ГСИ) в плоском пространстве. Мизнер, Руффини и др. рассмотрели ГСИ скалярных и тензорных волн в римановой геометрии; Соколов и др. [30] рассчитали поляризационные свойства ГСИ. Не имея еще возможности утверждать реальность синхротронной или иной природы всплесков Вебера, мы видим, что его эксперименты уже сыграли стимулирующую роль [1]).

§ 3. НОВЫЕ КАТАЛОГИ ТЕОРИЙ ГРАВИТАЦИИ

1. Каталог Торна

Многочисленные попытки обобщений ОТО до сих пор не были надлежащим образом классифицированы. Разумно различать, во-первых, разные попытки построения геометризованной электродинамики («единые» теории 20-х годов), связанные с отказом от римановой геометрии (Вейль, Эйнштейн, Эддингтон и др.). Мы не будем останавливаться на этом «первом» каталоге, который, как известно, не привел к успеху, хотя с нынешней, уже не столь полемической точки зрения мы видим, что от него осталось наследие в виде ряда полезных математических обобщений типа введения кручения и других.

В качестве «второго» каталога группа Торна недавно составила список «жизнеспособных» теорий, оценивая различные варианты главным образом с точки зрения удовлетворительного описания известных эффектов в рамках Солнечной системы. Наряду с ОТО в этом каталоге фигурируют прежде всего различные варианты СТТ (Йордана — Бранса — Дикке) и различные варианты конформно-плоских теорий. Вместе с тем в этом каталоге отсутствуют «нежизнеспособные» теории: кинематиче-

---

[1]) Улучшенная установка веберовского типа, по-видимому, не дала искомых всплесков (Тайсон, доклад в Нью-Йорке, 1972).

ский релятивизм Милна, творение материи Хойля и некоторые другие. Мы не считаем критерий подбора удовлетворительным, поскольку, например, ряд конформно-плоских теорий, включенных в каталог, также заведомо не удовлетворяет экспериментам. При всем том этот прагматический каталог заслуживает внимания, поскольку, в частности, все варианты характеризуются значениями некоторых параметров в разложении метрических компонент в первом постньютоновском приближении.

Речь идет только о «метрических» теориях, которые характеризуются: 1) наличием метрики $ds^2 = g_{\alpha\beta}dx^\alpha dx^\beta$, 2) уравнением для тензора энергии обычной материи $\nabla T = 0$, где ∇ — ковариантная производная. Идея метода восходит к Эддингтону, который предложил записать сферически-симметричный интервал в виде

$$ds^2 = \left(1 + \frac{b_1}{r} + \frac{b_2}{r^2} + \ldots\right) dt^2 - \left(1 + \frac{a_1}{r} + \frac{a_2}{r^2} + \ldots\right) dr^2 - r^2 d\Omega^2.$$

Ньютоновское приближение требует $b_1 = -2m$; отклонение света показывает, что $a_1 = b_1$. Группа Торна записывает метрику в форме

$$g_{00} = 1 - 2U + 2\beta U^2 - 4\Phi + \zeta a,$$

$$g_{0\alpha} = \frac{7}{2}\Delta_1 V_\alpha + \frac{1}{2}\Delta_2 W_\alpha,$$

$$g_{\alpha\beta} = -(1 + 2\gamma U)\,\delta_{\alpha\beta},$$

где U, Φ, V, a, W характеризуют распределение вещества, скорости и внутреннюю энергию, а 8 параметров γ, β, β_1, β_2, β_3, β_4, Δ_1, Δ_2 различны в случае ОТО, СТТ и в других вариантах ($\zeta = 0$). Не входя в детали, заметим лишь, что

$$U(x, t) = \int \frac{\rho(x', t)}{|x - x'|}\, dx',$$

$$\Phi(x, t) = \int \frac{\rho(x', t)\,\varphi(x', t)}{|x - x'|}\, dx',$$

$$\varphi = \beta_1 V^2 + \beta_2 U + \frac{1}{2}\beta_3 \Pi + \frac{3}{2}\beta_4 \frac{p}{\rho}.$$

Наглядно говоря, γ дает меру искривления пространственной метрики единичной массой, β дает меру нелинейности и т. д. (p — давление, Π — внутренняя энергия). Эксперименты дают значения параметров $\gamma = 1{,}04 \pm 0{,}08$, $\beta = 1{,}14 \pm 0{,}3$. (В нормировке Вилла параметры характеризуют сохранение энергии, момента и т. д.)

С этой точки зрения ОТО, удовлетворяющая с достаточной точностью всем известным данным, характеризуется всеми параметрами, равными единице. В ОТО можно еще ввести произвольный параметр — космологическую постоянную. Для СТТ имеем

$$\gamma = \frac{1+\omega}{2+\omega}, \quad \beta = 1 + \Lambda, \quad \beta_1 = \frac{3+2\omega}{4+2\omega} \quad \text{и т. д.},$$

где ω — параметр связи, Λ — космологическая постоянная.

После ОТО и СТТ в каталоге Торна идут конформно-плоские теории, в которых имеется глобальная лоренцева метрика $\eta_{\alpha\beta}$ и скалярное поле ψ, ведущее к физической метрике путем алгебраического соотношения $g_{\alpha\beta} = \psi\eta_{\alpha\beta}$, а само поле порождается согласно уравнению волнового типа $W(\psi) = 4\pi\rho$. Общая форма подобных теорий задается при помощи двух произвольных функций f и $\varkappa$:

$$g_{\alpha\beta} = \exp\{-2f(\Phi)\}\,\eta_{\alpha\beta}, \quad \eta^{\alpha\beta}\Phi_{,\alpha\beta} = 4\pi\varkappa(\varphi)\,\rho,$$
$$f(\varphi) = \varphi + q\varphi^2 + \ldots, \quad \varkappa(\varphi) = 1 + p\varphi + \ldots.$$

В постньютоновском приближении имеются два произвольных параметра p и q; $\gamma = -1$, $\beta = 1 - q$, $\beta_1 = 0$ и т. д. Хотя подобные схемы приводят к отсутствию запаздывания времени и отсутствию отклонения света в поле тяготения, все же анализировались следующие частные случаи:

$$\varkappa(\varphi) = \exp\{-4f(\varphi)\} \quad \text{(Уитроу — Мордух)},$$
$$f(\varphi) = \varphi, \quad \varkappa(\varphi) = e^{-4\varphi} \quad \text{(Нордстрем)},$$
$$f(\varphi) = -\log(1-\varphi), \quad \varkappa(\varphi) = (1-\varphi)^3 \quad \text{(Нордстрем, Эйнштейн — Фоккер)}.$$

Далее в каталоге Торна фигурируют теории «слоистого» типа, в которых только пространственная часть является конформно-плоской, допускающие тем самым наличие универсальной системы отсчета, скажем выделенной космологическим распределением материи. В основе лежит метрика

$$ds^2 = e^{2f(\varphi)}\,dt^2 - e^{2g(\varphi)}\,(dx^2 + dy^2 + dz^2).$$

Метрика фона является плоской, скалярное поле соответствует универсальной координате времени. Частным примером является $ds^2 = (1 - 2\varphi)\,dt^2 - (dx^2 + dy^2 + dz^2)$ (эйнштейновская теория переменной скорости света, которая, очевидно, резко не соответствует наличию отклонения света, запаздыванию времени, сдвигу перигелия). Другой вариант теорий с переменной скоростью света предложили Пейдж и Тупнер:

$$ds^2 = f^2(\varphi)\,(c^2\,dt^2 - dx^2 - dy^2 - dz^2).$$

Весьма простая схема этих авторов при подборе параметров a_1, a_2, $f = 1 + a_1\varphi + a_2\varphi^2 + \ldots$ согласуется с опытом. Ильмаз использует метрику

$$ds^2 = e^{-2\varphi}\,dt^2 - e^{2\varphi}(dx^2 + dy^2 + dz^2)$$

или

$$ds^2 = e^{\chi}\,dt^2 - e^{\varphi}(dx^2 + dy^2 + dz^2).$$

Розен исходит из метрики

$$ds^2 = \Phi\,dt^2 - \Psi^2(dx^2 + dy^2 + dz^2).$$

Общее заключение Торна — Ни гласит, что с основными опытами согласуются только ОТО, а также СТТ и слоисто-конформные теории. Для дальнейшего отбора теорий необходимы опыты и наблюдения высокой точности.

2. Скалярно-тензорная теория

Среди обобщений ОТО последнего времени наиболее интенсивно развивалась скалярно-тензорная теория (СТТ) Йордана — Бранса — Дикке, зародившаяся, как известно, на базе идеи Дирака о переменной гравитационной константе. Более последовательно было предложено рассматривать константу как скаляр, что привело к простейшему вариационному принципу

$$\delta S = \delta\int\left(\Phi R - L_m - \omega\,\frac{\Phi_{,\nu}\Phi^{,\nu}}{\Phi}\right)\sqrt{-g}\,d^4x = 0, \qquad (1)$$

где $\Phi^{-1}(x^\nu)$ интерпретируется как переменная гравитационная константа, ω — безразмерная константа связи. Вместе с тем СТТ является одним из простейших обобщений ОТО, учитывающих поля спина 0 и 2.

Для наглядности целесообразно рассмотреть нерелятивистское приближение, которое приводит к ньютоновским уравнениям

$$\Delta\varphi = 4\pi G\rho, \qquad (2)$$

$$\Delta f = -\frac{4\pi G}{\omega + 2}\rho \qquad (3)$$

(φ — ньютоновский гравитационный потенциал),

$$\Phi = c^4\chi\left[1 + \frac{1}{c^2}f(x^\nu)\right] \quad (\chi = \mathrm{const}). \qquad (4)$$

Важно подчеркнуть необходимость в данном приближении разложения вида (4), которое, в частности, не дает оснований делать заключения о возможности построения моделей звезд, качественно отличных от ньютоновских; устанавливать же нерелятивистские соотношения на базе иных произвольных соображений нецелесообразно.

Интересно рассмотреть квазиньютоновскую однородную космологическую модель в духе трактовки Милна — Мак-Кри в случае ОТО (следуя Горелику [49]). В этом случае для уравнений поля получаем

$$\chi \Delta\varphi = \frac{\omega+2}{2\omega+3}\rho + \frac{\omega}{\chi}\dot{\chi}^2 + \ddot{\chi},$$

$$\ddot{\chi} = \frac{\rho}{2\omega+3},$$

где $\chi(t) = \Phi/c^4$. Тогда уравнения космологической однородной модели фридмановского типа имеют вид

$$-3\chi\frac{\ddot{R}}{R} = \frac{\omega+2}{2\omega+3}\rho + \frac{\omega}{\chi}\dot{\chi}^2 + \ddot{\chi},$$

$$\ddot{\chi} = \frac{\rho}{2\omega+3},$$

где R — масштабный фактор. Не останавливаясь на анализе различных случаев, отметим, что при наиболее правдоподобном значении $|\omega| \gg 1$ уравнения дают:

1) $\chi \sim \chi_0 t$, где $\chi_0 = \mathrm{const}$; это соответствует первоначальной идее Дирака;

2) так как параметр ускорения $\ddot{R}/R$, согласно наблюдениям, отрицателен, получаем $\omega > 0$; Дикке на основе своеобразного принципа Маха также получает $\omega > 0$;

3) масштабный фактор меняется со временем по закону $R \sim t^{1/2}$.

Интересно подчеркнуть, что, как показал анализ Горелика точных решений СТТ, в случае вакуума и в случае наличия вещества существуют космологические однородные решения, в которых истинная сингулярность, связанная с обращением в бесконечность инварианта кривизны, наступает даже при конечных значениях масштабного фактора и параметра Хаббла. Напомним, что в противоположность этому в ОТО сингулярность в однородных космологических моделях наступает вместе с сингулярностью $(0, \infty)$ масштабного фактора и параметра Хаббла.

Обратимся теперь к коллапсу в рамках СТТ; дискуссия вокруг него приобрела добавочную актуальность ввиду резко различных полученных выводов. В противоположность Торну, считающему, что качественных различий между коллапсом в СТТ и ОТО нет, Саакян и Мнацаканян вообще отвергают неизбежность коллапса в СТТ, базируясь в известной мере на различной форме сферически-симметричных решений Гекмана — Йордана и Бранса — Дикке. Более детальное исследование [49] всей этой проблемы показало, что решения Гекмана и Бранса — Дикке по существу совпадают и переходят одно в другое путем преобразования сферических шварцшильдоподобных координат в изотропные. Применяя понятие сингулярной области как топологического пространства, удается показать, что сингулярная область является $(2+1)$-мерной, хотя в метрическом смысле при $\omega > 0$ она является точкой.

При исследовании движения пробных частиц в сферически-симметричном поле в СТТ показано, что при $\omega > 0$ частица достигает сингулярности за конечное координатное время (в отличие от ОТО) и, следовательно, звезда коллапсирует за конечное время. Действительно, ограничиваясь для иллюстрации случаем $\omega \gg 1$, получаем

$$8r_0 \frac{d\rho}{dt} = -\rho^{1-1/2\omega},$$

где $r_0 = 2mG/c^2$, а $r_0 + \rho$ — радиальная координата падающей частицы. Решение этого уравнения

$$\rho^{1/2\omega} = \frac{t_0 - t}{16 r_0 \omega}$$

показывает, что время достижения сингулярности конечно.

Оказывается, что в СТТ, где радиальные геодезические не являются полными:

$$\Delta s = \int_{R_0}^{r_0'} \sqrt{g_{00}\left(\frac{dt}{dr}\right)^2 - g_{11}}\, dr < \infty,$$

сингулярность является «голой», а не скрытой за горизонтом событий, как в известном решении Шварцшильда в ОТО. Эти любопытные выводы открывают новые возможности проверки СТТ наблюдением последних стадий эволюции коллапсирующих звезд. В СТТ, как и в случае Керра, снимается вырождение особых поверхностей ОТО.

Наконец, следует указать, что как в ОТО, так и в СТТ можно сформулировать критерий того, что центр сферически-симметричной конфигурации, описываемой соответствующим внутренним решением (давно известным в случае Шварцшильда для различных уравнений состояния), является истинной точкой в топологическом смысле:

$$\lim_{b \to 0} \int_{\infty}^{R_0} \frac{b \exp\left(\frac{\lambda}{2}\right) dr}{r^2 \left[a^2 \exp(-\nu) - 1 - b^2 r^{-2}\right]^{1/2}} \neq 0$$

для метрики вида

$$ds^2 = e^{\nu}\, dt^2 - e^{\lambda}\, dr^2 - r^2\, d\Omega^2,$$

где R_0 такое, что

$$[a^2 \exp(-\nu) - 1 - b^2 r^{-2}]\,|_{R_0} = 0, \quad a = \text{const}.$$

С этой точки зрения следует иметь ввиду, что результаты Саакяна и Мнацаканяна, которые не удовлетворяют этому критерию согласно указанным выше выводам, не могут служить аргументом против отсутствия коллапса в СТТ [48].

Дальнейшими обобщениями СТТ являются: 1) локализация константы связи, что вводит новую произвольную функцию $\omega(x^\mu)$, 2) включение космологической «постоянной», в дальнейшем также локализованной: $\Lambda(x^\mu)$ [последнее перекликается с членом порождения материи $C(x^\mu)$ Хойля]. Распоряжаясь произвольными параметрами СТТ, можно удовлетворить всем наблюдениям. В заключение вновь напомним многие любопытные следствия ослабления константы тяготения для геологии (типа расширения Земли) (Йордан, Иваненко — Селитов и др.).

3. «Третий» каталог теорий гравитации

Нетрудно дополнить «второй» каталог (Торна) рядом неучтенных вариантов, например линейной схемой Тоннела — Мавридес, дающей в отличие от ОТО несколько иные астрономические поправки, и другими. Однако мы предпочитаем более принципиальную точку зрения и предлагаем существенно дополнить «второй» каталог целым классом обобщений ОТО, подсказываемых современной теорией поля и частиц, причем речь может идти о классической или квантовой гравидинамике. Первым важным пунктом является применение не метрических коэффициентов, но тетрад, которые, как известно, оказались необходимыми для учета взаимодействия фермионов (электронов, протонов и т. д.), описываемых спинорным уравнением Дирака, с гравитационным полем (Фок — Иваненко). Тонкие проблемы трактовки спиноров в римановой геометрии затем рассматривали Шредингер, Картан, Инфельд, Ван дер Варден, Уилер и многие другие. Это фундаментальное обстоятельство подсказывает тетрадную «ревизию» ОТО; в частности, использование метода Палатини путем варьирования по h^a_μ $\left(g_{\mu\nu} = h^a_\mu h^a_\nu\right)$ и коэффициентам вращения Риччи Δ^{ij}_σ [42] наряду с установлением тех или иных дополнительных условий, например квазигармонических условий Родичева или условий Швингера или Меллера: $\partial/\partial x^\sigma \left(h h^\sigma_a\right) = 0$, затем построение гравидинамики, в которой h^σ_a а не $g_{\mu\nu}$ являются потенциалами (в модельном варианте тетрадной гравидинамики Тредер вместо неоднородных уравнений Эйнштейна берет за основу однородное уравнение, в котором воздействие материи на геометрию, т. е. на тетрады, имеет потенциальный характер). Главные результаты тетрадного подхода заключаются в установлении правдоподобного выражения для энергии гравитационного поля

$$t^{\mu}(\alpha)=\frac{c}{4\pi G}\left[2\Delta(\beta,\ \xi\tau)\,C(\xi\alpha,\ \beta)-\frac{1}{2}\delta(\alpha\tau)\,R\right]h^{\mu}(\tau),$$

где наряду с коэффициентами Риччи введены объекты неголономности C, связанные с ними равенством

$$\Delta_{\sigma\mu\nu}=(-C_{\sigma\mu\nu}-C_{\nu\sigma\mu}+C_{\mu\nu\sigma}).$$

Заметим еще изящное выражение для скаляра кривизны

$$R=\Delta(\xi,\ \alpha\beta)\,C(\alpha\beta,\ \xi).$$

Тетрадный вид энергии гравитационного поля обладает рядом преимуществ и подсказывает, в частности, новый критерий наличия гравитационного излучения Дозморова $t^{\mu}(4)\,t(4)_{\mu}=0$, которому, как оказалось, удовлетворяют все разумные точные волновые решения ОТО и который снова приводит к вырожденному типу II Петрова. Тетрадный формализм естественно приводит к сохраняющимся величинам типа углового момента, что было в сущности недоступно ОТО; при этом естественно вводится кручение, индуцируемое распределением спинового момента. Как видим, «тетрадная ревизия» ОТО уже привела к интересным результатам [40, 41, 44].

Вторым существенным подходом, подсказанным теорией поля, является «компенсационная» трактовка гравитации, предложенная Утиямой и независимо несколько позднее нами (Бродский, Иваненко, Соколик). Речь идет о рассмотрении параметров группы Пуанкаре не как глобальных величин, но как локализованных величин, являющихся функциями 4-точек. Тогда для сохранения ковариантности следует компенсировать члены, возникающие от производных этих параметров, ввести «компенсирующую» производную, оказывающуюся, вообще говоря, совпадающей с обычной ковариантной, и т. д. [43].

Тем самым гравитационное поле возникает как некоторое компенсирующее поле («калибровочное» в другой терминологии), притом в общем случае ведущее не только к искривлению, но и к кручению. Подобная процедура ранее позволила посмотреть на электромагнитное поле как на компенсирующее (при локализации фазы) или на векторные мезоны как на «компенсаторы» преобразований в изопространстве, ранее проводившихся с постоянной фазой (Сакураи). Компенсационный подход подсказывает также естественность добавки к основному скаляру его квадрата R^2, анализировавшегося с разных сторон еще в период геометризованных обобщений. Что касается кручения, то мы вновь обращаем внимание на теорему Родичева, связанную с добавлением нелинейного члена типа ψ^3 в дираковском уравнении (Гейзенберга — Иваненко) в случае параллельного переноса спинора в закрученном пространстве. Конкретные эффекты при простом выборе типа кручения рассмотрели недавно Траутман и Пономарев [50, 51].

Не останавливаясь на других подходах к гравитации, подсказываемых теориями спина 2, групповыми методами и другим арсеналом теории частиц, отметим, что примером подобных идей является предложение учета мезонов спина 2 в гравидинамике, рассматриваемое Саламом в его статье, помещенной в настоящем сборнике. Как мы видим, идеи теории поля и частиц уже позволили плодотворно проанализировать многие стороны гравитации.

§ 4. КВАРКИ В АСТРОФИЗИКЕ

Мы не будем здесь рассматривать астрофизические проблемы, а подчеркнем лишь вновь важную роль, которую могут играть кварки внутри сверхплотных звезд, в гипотетическом предзвездном веществе и в первичном космологическом котле (гипотеза Иваненко — Курдгелаидзе, которую развивали Паччини, Бербидж, Итох, де Саббата, Дэшен и др. [52—54]). Действительно, на основании аргументов, аналогичных предсказанию нейтронных и гиперонных звезд, тяжелые частицы — кварки (или, возможно, партоны) с необходимостью должны образовывать плазму в сердцевинах звезд при плотностях $\sim 10^{43}$ см$^{-3}$.

Возвращаясь вновь к нашей первоначальной идее — объяснить огромную светимость квазаров за счет значительной энергии, выделяемой при слиянии кварков в барионы, Дэшен [54] на основании правдоподобных ядерных реакций приходит к заключению, что соответствующая температура кварковой плазмы равна

$$T_b = 4\left(\frac{A}{30}\right)\left(\frac{3}{2}\,q\right)^4 10^6\,\mathrm{K}$$

($A = M_q/M_n$, q — заряд кварка) при светимости $\sim 10^{46}$ эрг/с, массе квазара $\sim 10^{10}\ M_\odot$ и радиусе $\sim 10^{17}$ см. Эффективность переработки кварков в фотоны оценивается примерно в 50%, в то время как для превращения протонов в α-частицы в известных ядерных циклах получается значение всего 0,7%.

Отметим еще полученные недавно результаты Максюкова [55], который вновь рассчитал равновесные конфигурации, учитывая дополнительно все известные барионные резононы вплоть до Ω^--частицы. В частном случае идеального барионного газа получаем

$$M = 0{,}290 M_\odot, \quad R = 7{,}65 \text{ км}$$

вместо значений $M = 0{,}324\ M_\odot$, $R = 11{,}1$ км, получаемых при учете лишь барионов октета; были произведены также расчеты параметров гипотетических кварковых звезд при помощи компьютера. Не останавливаясь на анализе множества предложенных уравнений состояния (Саакян и др., Гаррисон, Накано, Уилер и др., Хагедорн и др., Курдгелаидзе), отметим, что, используя правдоподобные термодинамические предположения, а также понятия бутстрепа и дуальности, Хагедорн устанавливает полуфеноменологическую форму спектра масс, в которой существенную роль играет предельная температура $T_0 = 160$ МэВ, что приводит, в частности, к разумным следствиям в теории неупругого рассеяния электронов на протонах. Отсюда возникает также асимптотическая форма уравнения состояния при больших плотностях [56]

$$P = P_0 + \rho_0 \ln \frac{\rho}{\rho_0}$$

($P_0 = 0{,}314 \cdot 10^{14}$ см, $\rho_0 = 1{,}253 \cdot 10^{14}$ г/см^3). Аналогичные соображения развиваются также для предзвездного вещества и первичного космологического котла, где опять-таки космология и астрофизика идут рука об руку с новейшей теорией элементарных частиц и субчастиц.

§ 5. НОВЫЕ ПОДХОДЫ К ЕДИНОЙ ТЕОРИИ

Учет гравитации в физике элементарных частиц, например в духе Салама, а также влияние компенсационно-групповых подходов на гравидинамику представляют собой уже значительный шаг на пути построения объединенной картины физической реальности, включающей в себя как элементарные частицы, так и гравитацию. Мы по-прежнему считаем правдоподобной гипотезу индуцирования асимметрий основного состояния (вакуума) квантовой теории поля космологическими асимметриями типа выделенной оси времени, преимущественной концентрации барионов и т. д. В частном случае наша гипотеза несколько в ином варианте развивается также Гейзенбергом — Дюрром. В связи с этим отметим предложения трактовки частиц как своего рода «полузамкнутых» вселенных.

Наиболее радикальное решение заключалось бы в построении единой теории на базе некоторого универсального спинора, подчиняющегося основному нелинейному уравнению типа Гейзенберга — Иваненко — Бродского. Включение гравитационных воздействий на подобный праспинор через тетрадные коэффициенты является следующим приближением; однако последний шаг заключался бы с этой точки зрения в построении самих тетрад или компонент метрики через праспинор.

Мы не будем подробно останавливаться на новых обнадеживающих результатах теории элементарных частиц, основанных на нелинейном спинорном уравнении, лучше всего с самого начала учитывающем также унитарную симметрию. Тогда, применяя усовершенствованный пропагатор Гейзенберга — Наумова, удается получить совсем неплохое согласие с опытом для масс адронов, даже для семейства декуплета вплоть до Ω^--частицы [58]. При помощи менее строгого, но быстрого метода слияния удается получить магнитные моменты адронов [59].

В последних работах Дюрр возвращается к первоначальным нелинейным уравнениям, содержащим массу, и предлагает явную форму конструкции тетрад через спиноры. По существу здесь мы снова имеем своеобразное применение основной идеи де Бройля о слиянии. Таким путем можно надеяться получить обобщенную гравидинамику, содержащую эйнштейновскую ОТО, и теорию мезонов спина 2.

Программа универсального спинора подсказывается в последнее время также под влиянием фундаментального характера бинарного кода в кибернетике и логике. Используя идеи бинарного кода, Финкельстейн приходит к варианту квантованного пространства-времени с дискретностью одного времени, тогда как у Снайдера, напротив, дискретно одно пространство, а в нашем первоначальном варианте использовалась простая дискретная решетка пространства-времени. С этими идеями перекликаются работы Вейцзекера [60], Пенроуза и высказывания Уилера. Следует надеяться на плодотворное объединение идей кибернетического и нелинейного спинора. Благодаря трансмутациям частиц-гравитонов подобный универсальный праспинор, возможно, наиболее приспособлен и для описания первичного состояния, в котором, по-видимому, нет нынешнего резкого различия обычной материи и пространства-времени гравитации [19а, 64, 65].

Резюмируя, можно сказать, что экспериментальные, астрофизические, космологические, а также квантовотеоретические аспекты современной гравидинамики не только вывели эту область на первый план, но и продолжают оказывать все более глубокое влияние на другие разделы всей современной картины мира.

Д. Иваненко

ЛИТЕРАТУРА

К § 1

1. *Guth E.* в книге Relativity (Proceedings Midwest Conference), ed. L. Witten, Plenum, New York, 1970.
2. *Tonnelat M. A.*, Histoire du principe de relativité, Flammarion, Paris, 1971.
3. *Treder H. J.*, Die Relativitaet ser Traegheit, Akad. Verlag, Berlin, DDR, 1972.
4. *Treder H. J.*, Gravitationstheorie und Aequivalenzprinzip, Berlin, DDR, 1971 (готовится русский перевод).
5. *Lichnerowicz A.*, GRG, **1**, 235 (1971).
6. *Seifert H. J.*, GRG, **1**, 247 (1971).
7. *Kronheimer E. H.*, GRG, **1**, 261 (1971).
8. *Schmidt B. G.*, GRG, **1**, 269 (1971).
9. *Pachner J.*, GRG, **1**, 281 (1971).
10. *Carter B.*, GRG, **1**, 349 (1971).
11. *Hawking S. W.*, GRG, **1**, 393 (1971).
12. *Misner C. W.*, в книге Relativity, ed. L. Witten, Plenum, New York, 1970.
13. *Hawking S. W.*, *Penrose R.*, Proc. Roy. Soc., **314**, 529 (1970).
14. *Penrose R.*, Rivista Nuovo Cimento, **1**, 252 (1969).
15. *Geroch R.*, Journ. Math. Phys., **9**, 450 (1968).
16. Современные проблемы гравитации (Доклады 2-й Советской гравитационной конференции), изд. ТГУ, Тбилиси, 1968.
17. Тезисы докладов 5-й Международной гравитационной конференции, изд. ТГУ, Тбилиси, 1968.
18. *Camesind M.*, GRG, **2**, 387 (1971).
19. *Белинский В. А.*, *Лифшиц Е. М.*, *Халатников И. М.*, Усп. физ. наук, **102**, 463 (1970).

19a. Тезисы докладов 3-й Советской гравитационной конференции 1972 г., изд. ЕГУ, Ереван, 1972.

К § 2

20. *Shapiro I. I.*, GRG, **3**, 135 (1972).
21. *Shapiro I. I.*, et al., Phys. Rev. Lett., **26**, 1132 (1971).
22. *Shapiro I. I.*, et al., Phys. Rev. Lett., **26**, 27 (1971).
23. *Sramek R. A.*, Astrophys. Journ. Lett., **167**, 55 (1971).
24. *Muhleman D. O.*, et al., Phys. Rev. Lett., **24**, 1377 (1970).
25. *Wheeler J. A.*, *Ruffini R.*, *Rees M.*, Black Holes, Gravitational Waves and Cosmology, Gordon and Breech, New York, 1972.
26. *Weber J.*, Phys. Rev. Lett., **24**, 276 (1970).
27. *Weber J.*, Phys. Rev. Lett., **25**, 180 (1970).
28. *Weber J.*, GRG, **3**, 59 (1972).
29. *Захаров В. Д.*, Гравитационные волны в теории тяготения Эйнштейна, изд-во «Наука», М., 1972.
30. *Соколов А. А.* и др., Phys. Lett., **42A**, 43 (1972).

30a. *Bertotti B.*, *Cavalere A.*, Nuovo Cimento, **2B**, 223 (1971).

31. *Kafka P.*, Nature, **226**, 436 (1970).
32. *O'Connell G. F., Surmelian G. L.*, Phys. Rev., **D4**, 286 (1971).
33. *O'Connell R. F.*, GRG, **3**, 123 (1972).
34. *Trautman A.*, GRG, **3**, 167 (1972).
35. *De Sabbata V.* et al., Nature, **232**, 53 (1971).

К § 3

36. *Thorne K. S., Will C. M.*, Astrophys. Journ., **163**, 595 (1971).
37. *Ni W. T.*, Astrophys. Journ. (1972).
38. *Nordtvedt K.*, Phys. Rev., **D3**, 1683 (1971).
39. *Will C. M.*, Astrophys. Journ., **169**, 141 (1971).
40. *Иваненко Д.*, Convegno s. Relativita Generale, Firenze, 1965; Acta Phys. Academ. Sci. Hungaricae (Budapest), **32**, 341 (1972); Тезисы ГIII, Ереван, 1972; «Tensor» (Japan), **25**, 161 (1972).
41. *Родиев В. И.*, Изв. вузов, физ. (Томск), № 1, 142 (1965).
42. *Фролов Б. Н.*, Вестник МГУ, № 5, 48 (1963); № 2, 56 (1964).
43. *Коноплева Н. П., Попов В. Н.*, Калибровочные поля, Атомиздат, 1972.
44. *Иваницкая О. С.*, Обобщенные преобразования Лоренца и их применение, изд-во «Наука и техника», Минск, 1969.
45. *Jordan P.*, Schwerkraft und Weltall, Vieweg, Braunschweig, 1955; Zs. Phys. **157**, 112 (1959).
46. *Дикке Р.*, в сборнике «Гравитация и относительность», М., 1965.
47. *Salmona A.*, Phys. Rev., **154**, 1218 (1967).
48. *Саакян Г. С., Мнацаканян М. А.*, Астрофизика, **5**, 169 (1969).
49. *Горелик Г. Е.*, Изв. вузов, физ. (Томск), № 1, 56 (1973).
50. *Пономарев В. Н.*, Вестник МГУ (1973).
51. *Пономарев В. Н.*, Bull. Ac. Polon. d. Sci., **19**, 732 (1971).

К § 4

52. *Иваненко Д., Курдгелаидзе Д. Ф.*, Астрофизика (Ереван), **1**, 479 (1965); Изв. вузов, физ. (Томск), № 8, 39 (1970); Lett. Nuovo Cimento, **2**, 13 (1969).
53. *De Sabbata V.* et al., Nuovo Cimento, **A45**, 513 (1966).
54. *Dashen* et al., Astrophys. Journ., **9**, 163 (1971).
55. *Максюков Н. И.*, Изв. вузов, физ. (Томск) (1973); Вестник МГУ (1973).
56. *Hagedorn R.*, Nuovo Cimento, **56A**, 1027 (1968).

К § 5

57. *Иваненко Д.*, Nuovo Cimento, **51A**, 244 (1967).
58. *Наумов А. И., Нгуен Н. Зао,* Вестник МГУ, № 2, 76 (1972); Ядерная физ., **7**, 664 (1968).
59. *Курдгелаидзе Д. Ф., Басьюни А.*, Ядерная физ., **9**, 432 (1959).
60. *Weizsaecker C.*, Zs. Naturforsch, **83a**, 245 (1958).
61. *Силенко П. Н.*, Изв. вузов, физ. (Томск), № 3, 48 (1971).
62. *Брагинский В. Б.*, Физические эксперименты с пробными телами, изд-во «Наука», М., 1970.
63. *Зельдович Я. Б., Новиков И. Д.*, Теория тяготения и эволюция звезд, изд-во «Наука», М., 1971.
64. *Wheeler J. A.*, Magic Whithout Magic, Collection of essays in honour of his sixth birthday, ed. J. R. Klauder, W. H. Freeman, San Francisco, 1972.
65. *Weinberg S.*, Gravitation and Cosmology, John Wiley, New York, 1972.

Статья 1

Сингулярности в общей теории относительности

Р. ГЕРОК

R. Geroch, в книге Relativity, Proc. Relativity Conf., Cincinnati, June 2—6, 1969, New York — London, 1970.

Цель настоящего обзора состоит, во-первых, в изложении в общих чертах сложившегося в настоящее время положения в исследовании проблемы сингулярностей в общей теории относительности (ОТО) и, во-вторых, в обсуждении некоторых наиболее важных нерешенных вопросов в этой области и оценке возможных путей их решения. Мы исходим из оптимистической точки зрения и концентрируем наше внимание в основном на тех результатах, которые мы желали бы получить, а не на тех результатах, которые, вероятно, будут получены в ближайшем будущем.

Нас будет особенно интересовать задача дальнего прицела: получение более общего подхода к сингулярностям в ОТО. В настоящее время существует большое число различным образом сгруппированных определений, теорем и понятий относительно сингулярностей, которые в конце концов необходимо объединить таким образом, чтобы они образовали некоторую цельную картину. В качестве первого шага в выполнении этой программы следует построить набор «стандартных» определений. Однако, по-видимому, почти хроническая трудность проблемы сингулярностей состоит в том, что в то время, как некоторая интуитивная идея формулируется в строгом виде, приходят в голову пять или десять новых возможностей. Хотя это явление, по-видимому, имеет место в каждой новой области знания, мы, вероятно, начинаем достигать того уровня, когда можно надеяться преуспеть в обобщении материала в виде немногих стандартных определений и достаточно простых теорем. После этого проблема сингулярностей в ОТО займет свое место в элементарных руководствах.

В трех параграфах этого обзора рассматриваются соответственно определение, существование и свойства сингулярностей в ОТО.

В § 1 обсуждается вопрос о формулировке подходящего определения сингулярного пространства-времени. Будет построена общая схема, в рамках которой могут быть сформулированы различные определения. Основная проблема заключается в выборе одной из двух возможностей: либо искать определение, пригодное для доказательств теорем (и при этом постоянно сталкиваться с трудностями, связанными с геодезической неполнотой), либо искать интуитивно более удовлетворительное определение (но в этом случае трудно остановиться на каком-либо определении). Мы надеемся, что в конце концов будет найдено определение, удовлетворяющее обоим критериям.

В § 2 формулируется следующий вопрос: каким образом сингулярные решения уравнений Эйнштейна распределены среди множества всех решений этих уравнений? Этот аспект проблемы в настоящее время сравнительно хорошо понят. Мы теперь в состоянии дать достаточно полный, хотя и не очень строгий ответ на поставленный вопрос. Было бы особенно интересно сделать этот ответ более строгим, а именно переформулировать теоремы Хокинга и Пенроуза [15—17, 19, 30, 32] таким образом, чтобы они говорили не о свойствах отдельных решений уравнений Эйнштейна, а о свойствах всего пространства решений в целом. Наиболее серьезная трудность в выполнении этой программы состоит в отсутствии подходящей топологии на множестве решений уравнений Эйнштейна.

В § 3 дается обзор различных попыток описания и классификации сингулярностей в ОТО. Это делается с целью определить множество «сингулярных точек», которое следует присоединить к пространственно-временному многообразию. Свойства сингулярностей тогда будут описываться локально через свойства этих дополнительных сингулярных точек. Из трех рассмотренных в этом параграфе подходов первый в принципе дает подробную информацию о сингулярных точках, но он слишком сложен для практического применения; второй подход достаточно прост, но нечувствителен к тонким деталям структуры сингулярностей; третий подход еще далек от того, чтобы быть надлежащим образом сформулированным.

Некоторые более конкретные (и более технические) вопросы изложены в приложениях. Первые три приложения описывают вспомогательный аппарат для работы с сингулярным пространством-временем: в них содержится определение «сингулярностей», основанное на поведении некоторых времениподобных кривых, понятия изометрического расширения пространства-времени и топологии для множества пространственно-временных многообразий. Четвертое приложение содержит упрощенное (и менее строгое) доказательство теоремы Хокинга [15, 17] о существовании сингулярностей. Это доказательство приведено здесь с целью дать почувствовать читателю основную нить доказательства и общую идею того, вследствие каких обстоятельств для справедливости теоремы требуется выполнение ряда дополнительных условий. Наконец, пятое приложение содержит список проблем, от общих формулировок того, что надлежало бы еще проделать в этой области, до точных утверждений, нуждающихся только в доказательстве или нахождении контрпримера. Каждое приложение, за исключением последнего, можно читать независимо от остальной части обзора.

§ 1. ОПРЕДЕЛЕНИЕ СИНГУЛЯРНОГО ПРОСТРАНСТВА-ВРЕМЕНИ

С нашей точки зрения, в настоящее время все еще не найдено вполне удовлетворительного определения сингулярного пространства-времени [1]). В идеале хотелось бы найти такое определение, которое было бы интуитивно удовлетворительным, простым для понимания и в то же время пригодным для доказательства теорем о существовании сингулярностей. Вполне возможно, разумеется, что мы хотим слишком многого: быть может, вообще не существует определения, которое удовлетворяло бы всем нашим критериям. С другой стороны, проблема нахождения удовлетворительного определения крайне важна, так как решение ее представило бы собой огромный шаг на пути к объединению различных разрозненных идей и к выяснению многих других вопросов, связанных с сингулярностями в ОТО.

Первый важный момент в проблеме сингулярностей заключается в том, что они не могут быть представлены как точки пространственно-временного многообразия [8, 26, 35]. Причина в основном состоит в том, что, задавая только «несингулярные» области пространства-времени, нельзя получить никаких сведений относительно количества и расположения дополнительных сингулярных точек, которые необходимо включить в общую картину. Эта ситуация, по-видимому, противоположна той, которая имеет место, скажем, в электродинамике, где основная минковская метрика позволяет определить все многообразие до того, как вводится в рассмотрение интересующая нас величина — тензор электромагнитного поля. В противоположность этому

---

[1]) Под пространством-временем мы будем понимать (связное, хаусдорфово) четырехмерное многообразие, имеющее (C^∞) метрику g_{ab} с сигнатурой $(+, -, -, -)$.

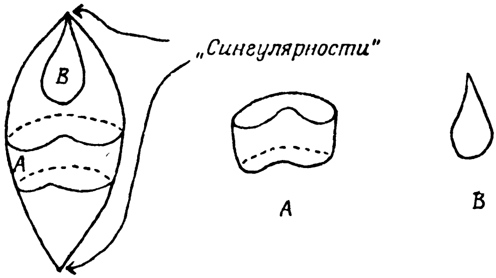
в ОТО «основная метрика» есть то самое поле, сингулярности которого мы хотим описывать. Строго говоря, следует говорить о «сингулярном пространстве-времени», а не о «сингулярных точках» в пространстве-времени.

В определении сингулярного пространства-времени должны быть соединены два различных понятия. Рассмотрим в качестве примера пространство-время M, состоящее из малой открытой окрестности одного из однородных пространственноподобных сечений модели Фридмана. Это пространство-время M в обычном интуитивном понимании не «сингулярно» (например, ни один из скалярных инвариантов не обращается в бесконечность). Тем не менее мы хотели бы исключить пространство-время M из рассмотрения в качестве модели вселенной на том основании, что M представляет собой только «часть вселенной». Сингулярности, которые, конечно, присутствуют в моделях Фридмана, не проявляют себя в M, так как пространство-время M можно (изометрически) расширить до полной модели Фридмана. Поэтому первый шаг состоит в том, чтобы научиться распознавать расширяемое пространство-время. Предположим теперь, что имеется пространство-время M', которое нерасширяемо (например, полная модель Фридмана). Сформулируем теперь второй критерий «сингулярности» пространства-времени M'. Важный момент заключается в том, что на данном этапе мы не надеемся обнаружить сингулярности в расширяемом пространстве-времени, так как сингулярности всегда возникают «на краю» пространственно-временного многообразия, а именно этот «край» может не содержаться в расширяемом пространстве. Поэтому теперь нам нужно только сформулировать определение «сингулярности», пригодное для нерасширяемого пространства-времени. Затем мы можем попытаться обобщить наше определение на расширяемые пространства. Таким образом, мы наметили три стадии в формулировке искомого определения: 1) определение «изометрической расширяемости»; 2) определение «сингулярности» для нерасширяемого пространства-времени и 3) определение «сингулярности» для любого пространства-времени. (Только вторая стадия содержит трудности.)

Легко сформулировать, что означает изометрическая расширяемость пространства-времени: пространство-время M *расширяемо*, если оно изометрично собственному подмножеству некоторого другого пространства-времени M', т. е. если M может быть «увеличено» как пространство-время. Можно рассматривать нерасширяемость пространства-времени как разумное физическое условие, которое необходимо наложить на любые модели вселенной. (Почему, в конце концов, природа должна остановить построение нашей вселенной на стадии пространства-времени M, когда она с тем же успехом могла бы продолжать до стадии пространства-времени M'?) Кроме того, расширяемое пространство-время создает довольно непривлекательную обстановку для некоторых наблюдателей, которые, хотя и движутся по геодезическим, тем не менее могут проводить свои эксперименты только конечное собственное время. Ряд вопросов относительно расширения пространства-времени рассматривается в приложении Б.

Следующая стадия состоит в формулировке определения «сингулярности», пригодного для нерасширяемого пространства-времени. Основным обстоятельством, которое может быть учтено при построении этого определения, является судьба надлежащим образом выбранных, находящихся в достаточно хороших условиях наблюдателей. Можно представить себе по крайней мере две опасности, которые могут подстерегать наших наблюдателей: либо они могут находиться на мировой линии, имеющей только конечную полную длину, либо они могут быть разорваны неумеренно большими приливными силами. Учет первой опасности приводит к определениям, согласно которым нерасширяемое пространство-время сингулярно в том случае, если существует времениподобная мировая линия, в определенном смысле удовлетворительная (например, соответствующая ограниченному ускорению), но имеющая конечную полную длину (см. [8, 23] и приложение А). Учет второй опасности приводит к определениям, использующим особенности поведения тензора Римана или некоторых скалярных инвариантов вдоль времениподобных кривых. Подобным образом легко сформулировать многочисленные точные критерии того, что представляет собой сингулярное нерасширяемое пространство-время; проблема заключается в том, чтобы решить, насколько серьезной должна стать опасность, прежде чем мы назовем пространство-время сингулярным, и чтобы выбрать относительные веса для обеих опасностей. Эта стадия — определение сингулярного нерасширяемого пространства-времени — представляет собой основу всей проблемы. К сожалению, в любом случае достаточно трудно видеть, каким образом определение, которое не использует существенно понятие геодезической неполноты, приведет к теоремам о существовании сингулярностей в ОТО.

Предположим теперь, что мы уже сформулировали определение «сингулярности», дающее разумный ответ для нерасширяемого пространства-времени. Тогда существует естественный путь обобщения этого определения на любое пространство-время. Пространство-время M будет называться сингулярным, если каждое его расширение M', которое само нерасширяемо, сингулярно (как нерасширяемое пространство-время). Согласно этому определению, каждое открытое подмножество несингулярного пространства-времени несингулярно. Например, произвольное открытое подмножество пространства Минковского, будучи геодезически неполным, не может считаться сингулярным.



Фиг. 1. Модель Фридмана [два пространственных измерения опущены: пространственноподобные сечения (3-сферы) изображаются на фигуре окружностями].

Открытые подмножества A и B определяют два различных пространства-времени. Хотя как A, так и B геодезически неполны, только B G-сингулярно.

В качестве иллюстрации этих замечаний рассмотрим, каким образом для наших целей подходит понятие геодезической неполноты. Для нерасширяемого пространства-времени мы отождествим понятия «сингулярности» и «времениподобной геодезической неполноты» (т. е. будем иметь дело только с временем жизни неускоренных наблюдателей). Окончательное определение тогда будет иметь следующий вид: пространство-время M называется *G-сингулярным*, если каждое его расширение времениподобно геодезически неполно. Таким образом, только та геодезическая неполнота, которая не может быть исключена при помощи расширения пространства-времени, может рассматриваться как признак сингулярности. Рассмотрим, например, модель Фридмана (фиг. 1). Пространство-время A определено как окрестность одного из однородных пространственноподобных сечений. Хотя A геодезически неполно, оно *не* G-сингулярно, так как, разумеется, существует расширение пространства-времени A (не модель Фридмана!), которое геодезически полно. С другой стороны, пространство-время B включает область, которая в пределе при $R = 0$ становится сингулярной. Каждое расширение пространства-времени B неполно, и поэтому само B является G-сингулярным. Полная модель Фридмана нерасширяема и неполна и поэтому G-сингулярна. Аналогично (расширенное) решение Шварцшильда [22], (расширенное) решение Керра [3], пространство Тауба [38] (т. е. область между границами Мизнера [26, 28]) G-сингулярны, в то время как пространство Минковского и плосковолновые решения [20] не G-сингулярны.

Но какое отношение все это имеет к теоремам о существовании сингулярностей — теоремам, в которых даже не упоминается о расширении пространства-времени? Обычное утверждение состоит в следующем: «Если пространство-время удовлетворяет определенным условиям, то оно геодезически неполно». «Геодезическая неполнота» тесно связана как с «расширяемостью», так и с «G-сингулярностью», так что неполнота дает приемлемую замену существованию сингулярности. Пусть имеется пространство-время M, удовлетворяющее «определенным условиям», фигурирующим в предыдущем утверждении. Тогда M должно быть геодезически неполным. Отсюда не следует, однако, что пространство-время M G-сингулярно, так как M может быть неполным только потому, что оно расширяемо. (Каждое расширяемое пространство-время геодезически неполно [26, 35].) Таким образом, имеется две возможности: 1) пространство-время M расширяемо (в этом случае M неудовлетворительно в качестве модели вселенной согласно принципу нерасширяемости) или 2) пространство-время M нерасширяемо (в этом случае M, будучи нерасширяемым и геодезически неполным, G-сингулярно). Мы можем выразить нашу теорему в полностью эквивалентной форме: «Если пространство-время M удовлетворяет определенным условиям и если M нерасширяемо, то M G-сингулярно». То обстоятельство, что понятие G-сингулярности эффективно сводится к гораздо более простому понятию геодезической неполноты, является одной из основных причин, благодаря которым оказывается возможным получить теоремы о существовании сингулярностей в ОТО.

§ 2. СУЩЕСТВОВАНИЕ СИНГУЛЯРНОСТЕЙ

Многие из работ, выполненных до сих пор по проблеме сингулярностей в ОТО, проводились с целью сформулировать условия, при которых пространство-время становится геодезически неполным. Следовательно, положение, при котором возникает G-сингулярность, является одним из наиболее разработанных аспектов проблемы. Имеющийся теперь набор теорем [15—17, 19, 30, 32] вместе с определенными физическими аргументами рисует широкую картину того, каким образом сингулярные решения уравнений Эйнштейна распределены среди всех решений этих уравнений, — картину, которую, по-видимому, значительно не изменят будущие результаты.

Изобразим множество всех решений уравнений Эйнштейна в виде множества точек на листе бумаги. (Под «решением» здесь понимается пространство-время, для которого симметричный тензор энергии удовлетворяет некоторым разумным неравенствам; см. приложение Г.) Теперь зачертим карандашом все точки этого листа бумаги, соответствующие «сингулярным решениям» в некотором подходящем смысле. Вопрос заключается в следующем: как будет выглядеть возникающая при этом фигура? Будут ли это немногие белые точки на черном фоне, однородный серый фон (какой интенсивности?) или чередующиеся черные и белые пятна? В настоящее время мы имеем общее представление об ответе на поставленный вопрос. Сначала мы рассмотрим этот ответ и попытаемся определить, почему мы чувствуем, что он в основном правильный. Позднее в этом параграфе мы рассмотрим подход, позволяющий объединить наши знания о существовании сингулярностей — доказать теоремы не об отдельном решении, а о существовании и свойствах диаграммы, изображающей все решения уравнений Эйнштейна.

На самом деле нам следовало бы изобразить две диаграммы — одну для «замкнутых моделей вселенной» и другую для «открытых моделей вселенной» [1]), так как в этих двух случаях положение, по-видимому, совершенно различно.

Рассмотрим сначала открытую модель. С одной стороны, известно существование ряда несингулярных открытых моделей: различные плосковолновые решения [20], электромагнитные решения Бертотти и Робинсона [2, 34], решение осциллирующего жидкого шара [37], цилиндрическое пылевое решение Маитры [24] и многие другие. Более того, разумно предположить, что если метрику некоторых из этих решений подвергнуть малому возмущению, то в результате будет получена также несингулярная открытая модель. Например, можно было бы, разумеется, ожидать существования точного решения, представляющего со-

---

[1]) Пока не ясно, каковы наиболее подходящие определения замкнутой и открытой моделей вселенной. Например, существует непрерывное пространство-время, которое может быть покрыто однопараметрическим семейством компактных пространственноподобных сечений и которое также может быть покрыто однопараметрическим семейством некомпактных пространственноподобных сечений. Для наших целей, однако, под *замкнутой* моделью вселенной мы будем подразумевать пространство-время, содержащее компактное пространственноподобное 3-мерное многообразие. (К сожалению, существует компактное пространство-время, которое не является «замкнутой» вселенной даже в указанном смысле.)

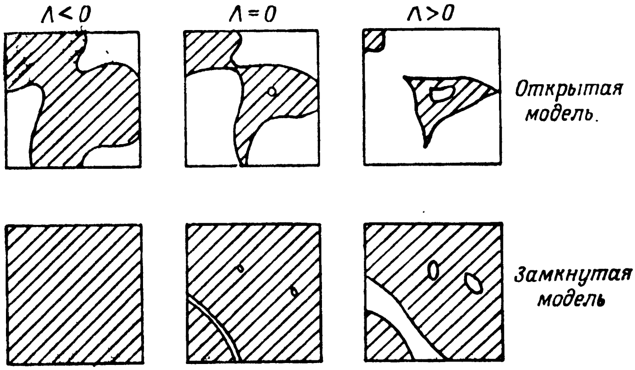
бой «почти минковское пространство», с небольшим количеством гравитационного излучения, приходящего из бесконечности, рассеивающегося и вновь уходящего в бесконечность. Подобные семейства решений будут изображаться на нашей диаграмме открытых моделей белыми пятнами.

С другой стороны, существуют теоремы о сингулярностях, пригодные для открытых моделей [15, 16, 19, 30, 32]. Однако все эти теоремы имеют характер «порога» развития сингулярности: если случится, что некоторое неравенство будет выполнено, то сингулярность будет развиваться (например, условие для ловушечной поверхности Пенроуза [30]). Подобные условия представляют собой не «разумные физические условия», накладываемые на модели вселенной, а скорее сигнал близкого коллапса. Следует, конечно, ожидать, что если некоторое решение удовлетворяет неравенству, сигнализирующему о начале коллапса, то любое пространство-время, получаемое достаточно малым возмущением такого решения, будет также удовлетворять этому неравенству. Подобные семейства сингулярных решений будут изображаться на нашей диаграмме черными пятнами.

Таким образом, можно ожидать, что наша диаграмма пространств открытых моделей будет содержать черные и белые пятна. В крайнем случае возможно, что диаграмма также будет содержать некоторые серые области, но это представляется менее вероятным.

Совершенно другую ситуацию мы имеем в случае замкнутых моделей вселенной: в этом случае теоремы несколько более строги. Имеют место теоремы [19], утверждающие, что замкнутая модель должна быть сингулярной тогда, когда она удовлетворяет условиям, которые, как можно ожидать, выполняются в «общем» случае, если только не нарушаются условия причинности. (Чувствуется, что эти условия причинности во многих случаях могут быть в конечном итоге опущены.) Теоремы для замкнутых моделей характерны отсутствием «пороговых» условий (хотя некоторые ими обладают; см., например, [17]). Таким образом, можно ожидать, что для замкнутых моделей диаграмма будет почти полностью черной. Однако будут наблюдаться по крайней мере несколько белых точек: существуют замкнутые геодезически полные плоские пространства-времена. (Класс подобных пространств-времен можно легко построить, используя компактное плоское трехмерное многообразие, предложенное Новацки [29].) Возможно, существует еще некоторое число других несингулярных замкнутых моделей, однако, как можно ожидать, они будут наблюдаться или как отдельные точки, или по крайней мере как области меньшей размерности на черном фоне диаграммы.

Во всех проведенных рассуждениях до сих пор неявно предполагалось, что космологическая константа Λ равна нулю. Чтобы нагляднее представить себе ситуацию, отметим, чтó произойдет, если это условие будет опущено. Грубо говоря, положительная космологическая константа имеет тенденцию создавать меньше сингулярных решений, в то время как отрицательная



Фиг. 2. Пространство решений уравнений Эйнштейна.

Шесть случаев представляют открытые и замкнутые модели для отрицательной, нулевой и положительной космологической постоянной. Заштрихованные («черные») области соответствуют решениям, сингулярным в некотором подходящем смысле.

космологическая константа дает противоположный эффект. Таким образом, в открытом случае увеличение Λ стягивает черные пятна и расширяет белые. В замкнутом случае положительная космологическая константа создает белые пятна (там, где ранее были только области меньшей размерности), в то время как при $\Lambda < 0$, как я полагаю, не существует даже одной несингулярной замкнутой модели. Таким образом, в замкнутом случае равенство $\Lambda = 0$ служит границей между существованием и несуществованием сингулярных решений [1]).

Сделанные замечания показывают, что мы пытаемся сформулировать теоремы, в которых делаются утверждения не о сингулярности отдельных решений уравнений Эйнштейна, а скорее о том, что диаграмма (фиг. 2), представляющая все решения

[1]) Очень трудно понять, как эти замечания относительно расширяющихся и стягивающихся пятен могут быть сделаны строгими, так как при изменении Λ пространство решений становится совершенно другим множеством. Другими словами, мы не можем указать на данное белое пятно на диаграмме при $\Lambda > 0$ и сказать, что это «то же самое» пятно, что и на диаграмме при $\Lambda = 0$, но только несколько увеличенное.

этих уравнений, обладает определенными свойствами. На пути реализации подобной программы возникает, разумеется, два препятствия: формулировка теоремы и ее доказательство. Но в действительности оказывается, что точная формулировка того, к чему нужно стремиться, — по крайней мере половина успеха. Отметим теперь один возможный подход к проблеме.

Пусть M — заданное (связное, хаусдорфово) четырехмерное многообразие. Обозначим через $\mathscr{D}(M)$ совокупность всех симметричных ковариантных тензорных полей второго ранга на M. Нас будут интересовать только те элементы множества $\mathscr{D}(M)$, которые могут быть физически интерпретированы как метрики для моделей вселенной. Поэтому введем подмножество $\tilde{\mathscr{D}}(M)$ множества $\mathscr{D}(M)$, состоящее из нерасширяемых метрик с сигнатурой $(+, -, -, -)$, тензор Риччи которых удовлетворяет энергетическому условию: величина $R_{ab}t^at^b$ неотрицательна для всех времениподобных векторов t^a (приложение Г). (Возможно, в определение множества $\tilde{\mathscr{D}}(M)$ нам следовало бы также включить условие причинности — например отсутствие замкнутых времениподобных кривых.) Наконец, обозначим через $\tilde{\mathscr{D}}_c(M)$ подмножество множества $\tilde{\mathscr{D}}(M)$, состоящее из замкнутых моделей, т. е. метрик, допускающих компактное пространственноподобное трехмерное подмногообразие, а через $\tilde{\mathscr{D}}_0(M)$ — подмножество открытых моделей (метрик, не допускающих подобного подмногообразия)[1]. Множества $\tilde{\mathscr{D}}_c(M)$ и $\tilde{\mathscr{D}}_0(M)$ представляют все замкнутые и все открытые модели, рассмотренные выше (фиг. 2, случай $\Lambda = 0$).

Следовало бы подчеркнуть еще один момент относительно пространства $\tilde{\mathscr{D}}(M)$. Пусть $g_{ab} \in \mathscr{D}(M)$ — лоренцева метрика на M, и пусть $g'_{ab} \neq g_{ab}$ — другая метрика, получаемая из g_{ab} диффеоморфизмом (гладким преобразованием) на M. Две метрики g_{ab} и g'_{ab} представляют в точности одну и ту же физическую модель, единственная разница между ними состоит в перенумерации точек многообразия M. Однако метрики g_{ab} и g'_{ab} определяют различные элементы множества $\tilde{\mathscr{D}}(M)$. Было бы, конечно, предпочтительнее, если бы различные элементы множества $\tilde{\mathscr{D}}(M)$ представляли физически различные модели, но какого-либо подходящего формализма для разрешения этой ситуации не было построено. К счастью, неоднозначность представле-

---

[1]) Напомним, что вопрос о том, замкнуто пространство-время или открыто, сильно зависит от метрики. Это не есть свойство исключительно самого исходного многообразия.

ния моделей множеством $\tilde{\mathscr{D}}(M)$ не умаляет интуитивного содержания теорем относительно $\tilde{\mathscr{D}}(M)$.

Элементы множеств $\tilde{\mathscr{D}}_c(M)$ и $\tilde{\mathscr{D}}_0(M)$ состоят из сингулярных и несингулярных пространств (для определенности G-сингулярных). Каким образом можно было бы описать тот факт, что множество $\tilde{\mathscr{D}}_0(M)$ содержит как сингулярные, так и несингулярные пятна? Одна возможность состоит в использовании языка топологии: можно было бы интерпретировать «пятно» как множество с внутренностью. Тогда мы могли бы поставить вопрос о том, какие, сингулярные или несингулярные, элементы множества $\tilde{\mathscr{D}}_0(M)$ образуют подмножества с непустой внутренностью. Тот факт, что во множестве $\tilde{\mathscr{D}}_0(M)$ не содержится серых областей, можно было бы интерпретировать как то, что каждый элемент множества $\tilde{\mathscr{D}}_0(M)$ принадлежит или одному из пятен, или границе одного из пятен. Другими словами, мы могли бы попытаться доказать, что каждый элемент множества $\tilde{\mathscr{D}}_0(M)$ принадлежит или замыканию внутренности множества сингулярных элементов, или замыканию внутренности множества несингулярных элементов. В случае замкнутых моделей нам хотелось бы доказать, что «общий» элемент множества $\tilde{\mathscr{D}}_c(M)$ сингулярен. На языке топологии можно интерпретировать свойство, выполняющееся для «общего» элемента, если оно выполняется на открытом и всюду плотном множестве. Тот факт, что множество сингулярных решений открыто, означает, что произвольное достаточно малое возмущение сингулярного решения вновь приводит к сингулярному решению; то, что это множество всюду плотно, означает, что независимо от малости разрешенного возмущения оно всегда переводит любое несингулярное решение в сингулярное.

Таким образом, мы были бы в состоянии по крайней мере сформулировать ряд предположений, если бы смогли задать подходящую топологию на множестве $\tilde{\mathscr{D}}(M)$. В приложении В обсуждаются свойства двух подобных топологий — «грубой» топологии C^p и «тонкой» топологии F^p. Которую из этих топологий нужно использовать, например, чтобы быть в состоянии утверждать, что множество сингулярных решений уравнений Эйнштейна открыто и всюду плотно на множестве $\tilde{\mathscr{D}}_c(M)$? Проблема заключается в том, что для этой цели не пригодна ни одна из этих двух топологий. Топология C^p настолько «груба» (содержит так мало открытых множеств), что в этой топологии для множества слишком легко быть всюду плотным, в то время как топология F^p настолько «тонка», что для множества слишком легко быть открытым. Один из ответов мог бы заключаться в доказательстве утверждения, что сингулярные пространства открыты в топологии C^p и всюду плотны в топологии F^p [1]). Существует другой путь для интерпретации этого предположения. Представим себе, что существует некоторая «правильная» топология, в которой установлено, что сингулярные замкнутые модели открыты и всюду плотны. Мы не знаем точно, какова эта топология, но мы определенно ожидаем, что она будет более «тонкой», чем топология C^p, и более «грубой», чем топология F^p. Доказательство того, что множество открыто в топологии C^p и всюду плотно в топологии F^p, полностью эквивалентно доказательству того, что это множество открыто и всюду плотно в *любой* топологии, более «тонкой», чем C^p, и более «грубой», чем F^p. Таким образом, можно было бы установить, что сингулярные решения определяют множество, открытое и всюду плотное в «правильной» топологии, хотя мы и не знаем в точности, что собой представляет эта топология.

Другой путь заключается в том, чтобы попытаться найти более подходящую топологию на множестве $\tilde{\mathscr{D}}(M)$. Общий метод построения топологий дан в приложении В. Мне кажется, важно остановиться на одной или, возможно, двух топологиях, с которыми постоянно работать, а не изобретать новые топологии для каждой новой теоремы.

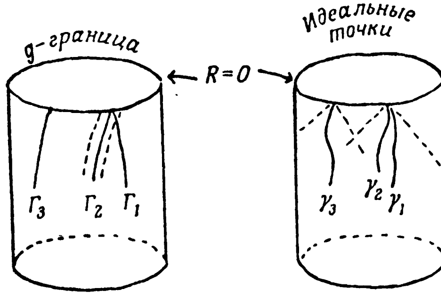
§ 3. СВОЙСТВА СИНГУЛЯРНОСТЕЙ

Установив, что большие классы решений уравнений Эйнштейна сингулярны, хотелось бы теперь лучше понять физику самих сингулярностей. Очень интересно было бы, например, классифицировать сингулярности, разделив их на три или четыре основных типа. Тогда можно было бы рассматривать более тонкие варианты теорем о существовании сингулярностей (определенные условия приводят к существованию сингулярностей определенных типов). Не преодоленная до сих пор трудность в классификации сингулярностей по их свойствам состоит в отсутствии вполне удовлетворительного способа обращения с сингулярным пространством-временем (скажем, G-сингулярным) с целью получения множества «сингулярных точек», локальные

[1]) Это утверждение, вероятно, неверно в той форме, в которой оно здесь сформулировано. Быть может, более разумно доказать, что некоторое подмножество множества сингулярных решений открыто (в топологии C^p) и всюду плотно (в топологии F^p). Например, подмножество плоскости, состоящее из точек (x, y), таких, что либо $x \neq 0$, либо $x = y = 0$, не открыто и не всюду плотно, но содержит открытое и всюду плотное подмножество.

свойства которых могут быть детально изучены. В этом параграфе мы кратко наметим три различных подхода к проблеме нахождения и классификации сингулярностей.

Первый подход [9, 10] исходит из сопоставления каждой неполной времениподобной геодезической воображаемой «конечной точке». Некоторые из этих точек затем отождествляются друг с другом, и совокупность подобных точек образует множество сингулярных точек. Грубо говоря, конечные точки, сопоставляемые геодезическим Γ_1 и Γ_2, отождествляются в том случае, если



Фиг. 3. Два способа анализа сингулярности в модели вселенной Фридмана (два пространственных измерения опущены).

На первой фигуре используется подход, основанный на понятии g-границы. «Конечная точка» геодезической Γ_2 отождествляется с «конечной точкой» геодезической Γ_1, но не геодезической Γ_3. Область между пунктирными линиями изображает множество точек, достигаемых геодезическими, в которые переходит геодезическая Γ_2 при малой вариации. g-граница представляет собой 3-сферу. На второй фигуре используется построение, основанное на понятии идеальных точек. Конечная точка, сопоставляемая кривой γ_1, отождествлена с конечной точкой кривой γ_2, но не кривой γ_3. Штриховые линии обозначают прошлое кривых. Множество идеальных точек также представляет собой 3-сферу.

геодезическая Γ_1 входит и остается в области, состоящей из всех точек, которые достигаются геодезическими, образующимися из геодезической Γ_2 малой вариацией (фиг. 3). На основе подобных отождествлений образуются отношения эквивалентности. Возникающее при этом множество классов эквивалентности конечных точек называют *g-границей.* Оказывается, что g-граница обладает весьма сложной структурой: можно обсуждать топологическую, причинную и в некоторых случаях даже метрическую структуру этого множества сингулярных точек. Именно в богатстве этой структуры и состоит одна из основных трудностей рассматриваемого подхода. Число мыслимых определений так велико, что весьма сложно решить, которые из них наиболее удовлетворительны. Далее, в этом подходе излишне усложнено развитие теории, так что трудно понять, как можно доказывать изящные теоремы при помощи столь громоздкого аппарата. С другой стороны, преимущество подхода, использующего понятие g-границы, состоит в том, что этот подход непосредственно связан с определением сингулярного пространства-времени в терминах геодезической неполноты. Таким образом, имеется по крайней мере возможность того, что некоторые теоремы о существовании сингулярностей могли бы быть переинтерпретированы как теоремы о структуре g-границы.

Второй подход [13] основан на сопоставлении любой непродолжаемой времениподобной кривой (не обязательно геодезической) воображаемой «конечной точке». Конечные точки, сопоставляемые кривым γ_1 и γ_2, отождествляются при условии, что прошлое[1]) кривой γ_1 совпадает с прошлым кривой γ_2 (фиг. 3). Соответствующие классы эквивалентности называют *идеальными точками*. Эти идеальные точки затем должны быть разделены на два класса, а именно на те, которые представляют собой «истинные сингулярные точки», и на те, которые представляют собой «точки на бесконечности». (В пространстве Минковского, например, идеальные точки определяют нулевую поверхность Пенроуза на бесконечности будущего [31].) Оба типа идеальных точек входят в рассмотрение на равных основаниях, так как полное сконструированное множество идеальных точек конформно инвариантно. С другой стороны, различие между точками на бесконечности и сингулярными точками не будет конформно инвариантным. До сих пор не найдено точной формулировки этого различия. Если отвлечься от этой последней проблемы, то приведенная выше характеристика сингулярных точек привлекает своей простотой. К сожалению, множество идеальных точек обладает слишком бедной структурой. (На этом множестве не найдено даже разумной топологии.) Далее, понятие идеальных точек в некоторых случаях не может помочь установить различие между конструкциями, которые можно было бы рассматривать как различные сингулярные точки. Мизнер [27] рассмотрел некоторые типы пространства-времени, содержащие только одну идеальную точку! (Сравните с моделью Фридмана, фиг. 3.)

В качестве примера указанных двух подходов рассмотрим замкнутую модель Фридмана (фиг. 3). В каждом случае множество сингулярных точек состоит из трехмерной сферы (а не точки!), присоединенной к пространству-времени при $R = 0$ (R — радиус сферического пространственноподобного сечения).

---

[1]) *Прошлое* некоторого множества определяется как совокупность всех точек, которые могут быть достигнуты из данного множества по времениподобной кривой, направленной в прошлое. (Предполагается, что на пространстве-времени задана временная ориентация и оно удовлетворяет условиям причинности.)

В случае построения, основанного на понятии g-границы, определяется (с некоторыми трудностями) топология, дифференцируемые структуры и метрика на множестве сингулярных точек. В случае построения, использующего понятие идеальных точек, множество сингулярных точек может быть найдено дополнительной проверкой, но, хотя, конечно, известно, какую топологию хотелось бы иметь на этом множестве, не существует никаких предписаний, которые были бы разумны в общем случае и которые давали бы желаемую топологию на фиг. 3.

Третий подход к классификации сингулярностей может быть основан на теоремах о погружениях. Как недавно показал Кларк [5], любое пространство-время может быть изометрично погружено в плоское пространство сигнатуры $(+, +, +, -, \ldots, -)$ (всего 87 минусов!). Идея третьего подхода заключается в сопоставлении пространству-времени M пространства с «присоединенными сингулярными точками» при помощи рассмотрения замыкания пространства-времени M в объемлющем пространстве. Конечно, при такой высокой размерности пространства нет никакой надежды на то, что погружение будет единственным. Однако, возможно, удастся доказать, что замыкание пространства-времени M существенно единственно или что определенные свойства замыкания независимы от выбора способа погружения.

Все три подхода обладают той общей чертой, что они пытаются связать сингулярные точки с произвольным пространством-временем. Представляется весьма вероятным, что для пространства-времени, удовлетворяющего энергетическому условию (приложение Г), классификация сингулярных точек явилась бы гораздо более простой задачей, чем в общем случае. Было бы интересно найти метод определения сингулярных точек, использующий некоторым существенным образом тот факт, что следует рассматривать только те типы пространства-времени, которые удовлетворяют энергетическому условию.

Наконец, не составило бы большого труда уточнить определенную выше классификацию сингулярных точек в следующих направлениях. (Для простоты рассмотрим только g-границу.) Пространство-время, состоящее из малого открытого подмножества пространства Минковского, не является G-сингулярным, как мы видели в § 1, тем не менее оно имеет нетривиальную g-границу. Было бы приятнее, однако, если бы можно было переопределить g-границу таким образом, чтобы она была пуста всегда, когда пространство-время не является G-сингулярным. Поэтому мы могли бы продолжить рассуждения следующим образом. Пусть M — произвольное пространство-время, и пусть M' — не расширяемое далее расширение пространства-времени M (которое существует; см. приложение Б). Те точки g-границы пространства-времени M', которые не находятся «вблизи» пространства-времени M, не могут рассматриваться как сингулярные точки пространства-времени M. Поэтому рассмотрим множество $\bar{M}$, объединяющее пространство-время M с множеством тех точек g-границы пространства-времени M', которые одновременно принадлежат замыканию пространства-времени M. Это множество $\bar{M}$, вообще говоря, не будет независимым от способа расширения M до M'. (Полезным упражнением было бы найти подходящий пример; выбирайте в качестве пространства-времени M произвольное открытое собственное подмножество пространства Минковского.) Однако, возможно, удастся найти в некотором смысле «минимальное» множество $\bar{M}$, образуя «пересечения» всех множеств $\bar{M}'$, получаемых подобным способом. На фиг. 1, например, «уточненная g-граница» пространства-времени A была бы пуста, в то время как подобная же конструкция для пространства-времени B состояла бы из одной или более точек, расположенных на остром конце.

Мне хотелось бы выразить благодарность Уолкеру и Пенроузу, прочитавшим рукопись и предложившим многочисленные улучшения.

ПРИЛОЖЕНИЕ А. T-ПОЛНОТА

Понятие T-полноты было введено Эресманом ([7], см. также [1]) и затем использовалось, особенно Аве [1], как гипотеза, принимаемая при доказательстве различных результатов о глобальных свойствах пространства-времени. Понятие T-полноты тесно связано с судьбой соответствующим образом определенных наблюдателей. В настоящем приложении мы дадим несколько отличную формулировку понятия T-полноты с целью использовать это рассмотрение для более удовлетворительного определения сингулярного пространства-времени.

Пусть M — пространство-время, и пусть $\gamma(s)$ — времениподобная кривая, параметризованная при помощи своей собственной длины s, отмеряемой от начальной точки p, находящейся в прошлом. Пусть ξ^a — единичный вектор, касательный к кривой γ. Выберем вдоль кривой γ три единичных вектора η^a_α ($\alpha = 1, 2, 3$), ортогональных друг другу и вектору ξ^a и переносимых вдоль кривой γ при помощи переноса Ферми:

$$\xi^b \nabla_b \eta^a_\alpha = -\xi^a \left(A_b \eta^b_\alpha \right),$$

где $A_b = \xi^a \nabla_a \xi_b$ — ускорение вдоль кривой γ. Компоненты ускорения относительно векторов η^a_α

$$A_\alpha(s) = A_b \eta^b_\alpha$$

мы будем называть *функциями ускорения* вдоль кривой γ. (Отметим, что, поскольку $A_b \xi^b = 0$, компоненты $A_\alpha(s)$ и векторы η^a_α определяют вектор ускорения A_b единственным образом.) Функции ускорения содержат полную информацию, требуемую, например, пилоту ракеты, желающему вести свой корабль строго вдоль геодезической линии γ.

Ясно, что заданные в точке пространства-времени M ортонормированные тетрады $\left(\xi^a, \eta^a_\alpha\right)$ совместно с набором $A_\alpha(s)$ функций ускорения определяют единственным образом некоторую кривую в пространстве-времени M. Однако эта кривая не обязательно должна быть определена для всех тех значений параметра s, для которых были первоначально заданы функции $A_\alpha(s)$, так как она может в некоторый момент времени «выйти за край» пространства-времени M.

Говорят, что пространство-время *T-полно* [1]), если для каждого набора функций ускорения $A_\alpha(s)$, задающих времениподобную кривую в пространстве Минковского, и для любого набора ортонормированных тетрад в точке пространства-времени M соответствующая кривая в пространстве-времени M определена для тех же значений параметра s, что и связанная с ней кривая в пространстве Минковского. (Интуитивно T-полнота означает, что пилот ракеты, ведущий свой корабль в пространстве-времени M так, как если бы он находился в пространстве Минковского и не выходил бы за край пространства Минковского [2]), не выйдет также за край пространства-времени M.)

Приведенное выше определение обладает двумя неприятными чертами: во-первых, на практике определить, является или нет данное пространство-время T-полным, оказывается достаточно сложной задачей (включающей исследование всех функций ускорения, связанных со всеми кривыми в пространстве

---

[1]) В обычном определении t-полноты допускаются также нулевые кривые. Мы этого здесь не делаем, так как это нарушило бы простоту нашей характеристики кривых на основе функций ускорения, и также потому, что, по-видимому, с физической точки зрения более разумно рассматривать мировые линии наблюдателей, а не фотонов.

[2]) Для построения времениподобной кривой, которая выходит за край пространства Минковского, нарисуйте кривую, уходящую в нулевую бесконечность [31] таким образом, что ее касательный вектор становится светоподобным вблизи бесконечности столь быстро, что эта кривая имеет конечную полную длину.

Минковского) и, во-вторых, приведенное определение решающим образом зависит от пространства Минковского как модели полного пространства-времени. Почему бы не выбрать для этой цели некоторые другие типы пространства-времени, например такие, как пространство де Ситтера? Оба эти возражения будут вскоре отвергнуты переформулировкой нашего определения.

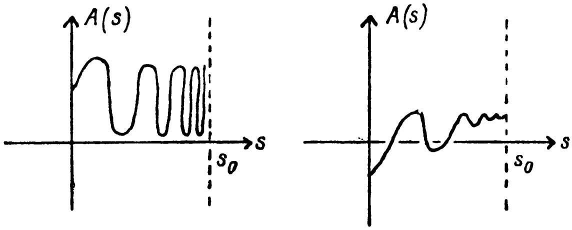
Поскольку нас интересует, что́ произойдет с наблюдателями после каждого конечного интервала времени, естественно обратить внимание на функции ускорения, определенные только для конечного изменения параметра s. Задавая набор функций ускорения $A_\alpha(s)$ класса C^∞, определенных на интервале $S \in [0, s_0]$, мы будем говорить, что функции ускорения $A_\alpha(s)$ *приспособлены* к пространству-времени M, если для каждой ортонормированной тетрады в точке пространства-времени M соответствующая кривая в M может быть продолжена за значение параметра $s = s_0$. (Другими словами, нашему пилоту ракеты, ведущему свой корабль в пространстве-времени M, мы можем совершенно безопасно задавать любые наборы функций ускорения, приспособленных к M.) Решающий фактор, определяющий, будет или нет данный набор функций ускорения $A_\alpha(s)$ приспособлен к пространству-времени M, заключается в поведении этих функций вблизи значения $s = s_0$. Действительно, можно сформулировать много различных определений сингулярного пространства-времени, отвечая на вопрос, приспособлены или нет к данному пространству-времени все функции ускорения, обладающие определенными свойствами вблизи значения $s = s_0$. Например, пространство-время будет полным в смысле ограниченности ускорения [23], если к нему приспособлены функции ускорения, ограниченные в окрестности значения $s = s_0$. Аналогично пространство-время будет (времениподобно) геодезически полным, если к нему приспособлены функции с нулевым ускорением.

Мы утверждаем, что пространство-время T-полно тогда и только тогда, когда к нему приспособлен любой набор функций ускорения, имеющих предельные значения при $s = s_0$ и принадлежащих к классу C^∞ в окрестности значения s_0. (Доказательство этого утверждения очень просто.) Заметим, что в приведенной формулировке даже не упоминалось пространство Минковского. Отсюда можно заключить, что в точности то же самое понятие T-полноты вытекало бы и в том случае, если бы в первоначальном определении термин «пространство Минковского» был заменен на любое другое пространство-время (даже на неполное пространство-время)!

Предыдущие замечания наводят на мысль ввести следующие определения: пространство-время T_n-*полно* ($n = 0, 1, \ldots, \infty$), если к нему приспособлен любой набор функций ускорения, имеющих предельное значение при $s = s_0$ и относящихся в окрестности значения s_0 к классу C^n. Таким образом, понятие T_∞-полноты эквивалентно понятию T-полноты. Введенные определения связаны следующей системой импликаций:

$$BA \Rightarrow T_0 \Rightarrow T_1 \Rightarrow \ldots \Rightarrow T_\infty \Rightarrow G, \qquad \text{(A.1)}$$

где BA обозначает полноту в смысле ограниченности ускорения, а G — (времениподобную) геодезическую полноту. (Полнота в смысле Мизнера [26, 35] заняла бы в (A. 1) место правее символа геодезической полноты.)



Фиг. A1. Типичные функции ускорения, используемые для определения того, является ли пространство-время BA-полным (левый график) или T_0-полным (правый график).

Для иллюстрации физического смысла понятия T_n-полноты сравним BA-полноту и T_0-полноту. Существует пример [8] пространства-времени, геодезически полного, но не BA-полного. Функции ускорения неполной кривой в этом примере показаны на фиг. A1. Мы видим, что BA-полнота с физической точки зрения представляется несколько более сильным требованием, налагаемым на пространство-время, так как ни один пилот ракеты не смог бы, по-видимому, провести свою ракету с ускорениями, показанными на левом графике фиг. A1. С другой стороны, понятие T_0-полноты связано с судьбой пилота ракеты, функции ускорения которой изображены на правом графике фиг. A1. Ясно, что они представляют собой для нашего пилота более приемлемую задачу. С физической точки зрения T_0-полнота, по-видимому, более интересна, чем BA-полнота.

Наконец, можно задаться вопросом, какие из импликаций (A.1) могут быть обращены. Единственный известный результат в этой области состоит в том, что из G-полноты не вытекает BA-полнота. Вероятно, справедливо утверждение, что ни одна из импликаций не обратима.

ПРИЛОЖЕНИЕ Б. РАСШИРЕНИЯ ПРОСТРАНСТВ

Мы разовьем формализм, позволяющий исследовать расширения различных типов пространства-времени. Этот формализм приводит не только к очень четкой формулировке двух фундаментальных конструкций, связанных с понятием расширения, но также к выяснению ряда важных и нерешенных проблем.

Пусть M и M' — два (связных) пространства-времени, и пусть символическая запись $M \leqslant M'$ означает, что пространство-время M' является (изометрическим) расширением пространства-времени M (т. е. что существует изометрия пространства-времени M на некоторое подмножество пространства-времени M'). Определяемое подобным образом соотношение оказывается, однако, не очень полезным, так как оно не определяет даже отношения частичного упорядочения (т. е. не выполняются такие аксиомы, как $M \leqslant M'$, $M' \leqslant M \Leftrightarrow M = M'$ и $M \leqslant M' \leqslant M'' \Rightarrow M \leqslant M''$) на совокупности всех пространств-времен [1]. Причина заключается в том, что совокупность всех пространств-времен не является самым удобным множеством с точки зрения проблемы расширения. Введем понятие *пространства-времени с системой отсчета*, означающее связное пространство-время M совместно с присоединенной в некоторой точке этого пространства-времени ортонормированной тетрадой. Пусть $\mathcal{J}$ обозначает совокупность всех пространств-времен с системой отсчета [2]. Символическая запись $M \leqslant M'$, где $M, M' \in \mathcal{J}$, означает существование изометрии пространства-времени M на подмножество пространства-времени M', причем эта изометрия переводит выделенную тетраду в пространстве-времени M в выделенную тетраду в пространстве-времени M'. Причина введения систем отсчета заключается в том, что если подобная изометрия существует, то она единственна [12]; таким образом, символ $\leqslant$ обозначает отношение частичного упорядочения на множестве $\mathcal{J}$. Хотя множество $\mathcal{J}$, разумеется, может быть наделено какой-либо еще структурой, кроме рассматриваемой структуры частичного упорядочения, оказывается, что бóльшая часть интересных вопросов, связанных с понятием расширения, может быть сформулирована как свойства отношения $\leqslant$.

---

[1] Например, пусть M обозначает подмножество $\{(t, x, y, z) \mid t < 0\}$ и M' — подмножество $\{(t, x, y, z) \mid t < 0, \quad t < x^2 + y^2 + z^2 - 1\}$ пространства Минковского. Тогда $M \leqslant M'$ и $M' \leqslant M$, но M и M' не равны (не изометричны).

[2] Именно на множестве $\mathcal{J}$ определено понятие предела [12]. Действительно, предельные переходы индуцируют топологию на множестве $\mathcal{J}$, очень схожую с C^∞-топологией (приложение В).

Подмножество (не пустое) $\mathscr{P} \subset \mathscr{J}$ будем называть *линейно направленным*, если для любых M, $M' \in \mathscr{P}$ существует такое $M'' \in \mathscr{P}$, что $M \leqslant M''$ и $M' \leqslant M''$. Например, если задано любое $M \in \mathscr{J}$, то можно образовать множество

$$I^{-}(M) = \{M' \mid M' \leqslant M\},$$

которое будет линейно направленным. (Это множество $I^{-}(M)$ можно интерпретировать как совокупность всех открытых связных подмножеств пространства-времени M, содержащих выделенную тетраду.)

Два очень полезных результата, связанных с понятием расширения, теперь могут быть выражены следующим образом.

Теорема Б1 [1]). Пусть множество $\mathscr{P}$ линейно направлено. Тогда существует единственное пространство-время с системой отсчета M, такое, что: 1) $\mathscr{P} \subset I^{-}(M)$ и 2) M минимально по отношению к свойству 1, а именно если $\mathscr{P} \subset I^{-}(M)$, то $M \leqslant N$.

Теорема Б2. Пусть $M, M' \in \mathscr{J}$, и допустим, что некоторый элемент множества $\mathscr{J}$ предшествует как M, так и M'. Тогда существует единственное пространство-время с системой отсчета M'', такое, что: 1) M'' предшествует как M, так и M' и 2) M'' максимально по отношению к свойству 1, а именно если $N \leqslant M$ и $N \leqslant M'$, то $N \leqslant M''$.

Первая теорема описывает операцию «сшивания» совокупности пространств-времени с образованием некоторого рода «объединения» с той особенностью, что некоторые изометричные области пространств-времен будут отождествлены. Вторая теорема утверждает, что если два пространства-времени согласованы друг с другом в некоторых окрестностях своих выделенных тетрад, то существует единственная максимальная окрестность, в которой они согласованы. Основные идеи доказательства обеих теорем изложены в работе [4] (хотя и не в столь общей форме, как здесь). В частности, из теоремы Б1 и леммы Цорна [21] непосредственно вытекает, что каждое пространство-время с системой отсчета содержится в нерасширяемом пространстве-времени (вообще говоря, не единственном).

С каждым свойством P пространства-времени, например со свойством быть «геодизически полным» или свойством «обла-

---

[1]) Математическая конструкция, аналогичная той, которая имеет место в теореме Б1, хорошо известна во многих областях математики, например в теории групп и в теории топологических пространств. Совокупность пространств-времен с системой отсчета образует категорию. (Морфизмами служат изометрии, сохраняющие выделенную тетраду.) Теорема Б1 утверждает, что в этой категории существуют прямые пределы [36, стр. 18]. Обратные пределы, однако, не существуют. Вместо этого имеет место теорема Б2.

дать двумя векторами Киллинга», свяжем подмножество $\mathcal{J}(P)$ множества $\mathcal{J}$, состоящее из всех тех пространств-времен с системой отсчета, которые обладают данным свойством P. Действительно, понятие «подмножество множества $\mathcal{J}$» означает в точности то же самое, что и понятие «свойство совокупности пространств-времен». Естественно задать вопрос о связи свойства P со структурой частичного упорядочения на множестве $\mathcal{J}(P)$. Особый интерес вызывают следующие три условия, выполняемые на множестве $\mathcal{J}(P)$:

1. Если $M' \in \mathcal{J}(P)$ и $M \leqslant M'$, то $M \in \mathcal{J}(P)$.

2. Если задано любое линейно направленное подмножество множества $\mathcal{J}(P)$, то единственный элемент M, определяемый теоремой Б1, также содержится в множестве $\mathcal{J}(P)$.

3. Если M — максимальный элемент множества $\mathcal{J}(P)$ (т. е. если из $M \leqslant M' \in \mathcal{J}(P)$ вытекает $M = M'$), то M будет также максимальным элементом множества $\mathcal{J}$ (т. е. из $M \leqslant N \in \mathcal{J}$ будет вытекать $M = N$).

Условие 1 означает, что каждое связное открытое подмножество пространства-времени, обладающего свойством P, также обладает свойством P. Условие 2 означает, что операция «сшивания», примененная к пространствам-временам, обладающим свойством P, в свою очередь приводит к пространству-времени, обладающему свойством P. Условие 3 означает, что обладающее свойством P пространство-время, которое не может быть расширено до более обширного пространства-времени, обладающего свойством P, не может быть расширено *вообще ни до какого* более обширного пространства-времени.

Обычно нетрудно выяснить, удовлетворяет или нет данное свойство условиям 1 или 2. Например, свойства «быть свободным от источников решением уравнений Эйнштейна», «иметь по крайней мере n векторов Киллинга» и «обладать везде тензором Вейля типа...» все удовлетворяют первым двум условиям. [Действительно, утверждение, что некоторое свойство удовлетворяет условиям 1 и 2, представляет собой хорошее определение того, что мы подразумеваем под «локальными» свойствами пространства-времени.] С другой стороны, условие 1 не выполняется для свойства «быть асимптотически простым» и свойства «обладать поверхностью Коши», в то время как условие 2 не выполняется для свойства «быть геодезически неполным» (все только что перечисленные свойства являются глобальными).

Более трудная — и более интересная — проблема заключается в выяснении того, какие свойства удовлетворяют условию 3. Например, для того чтобы определить, удовлетворяет ли или нет условию 3 свойство «быть свободным от источников решением уравнений Эйнштейна», необходимо решить, справедливо или нет следующее утверждение: если свободное от источников решение не может быть далее расширено как свободное от источников решение, то оно вообще не может быть далее расширено как пространство-время. В то время как это утверждение, по всей видимости, справедливо, не известно ни одного его доказательства. Предположим, что рассматриваемое утверждение ошибочно, например, для некоторого пространства-времени M с системой отсчета. В § 1 мы рассмотрели возможность введения понятия нерасширяемости как некоторого нового условия, которое необходимо наложить на модели пространства-времени. Давайте передадим наше пространство-время M природе и попросим ее превратить его в модель вселенной. Чтобы проверить выполнение принципа нерасширяемости, она должна попытаться расширить пространство-время M. Но M не может быть расширено без нарушения уравнений Эйнштейна. С целью избежать наложения слишком строгих ограничений на класс разрешенных моделей вселенной может возникнуть искушение подвергнуть принцип нерасширяемости ревизии — наложить на модели вселенной только требование невозможности дальнейшего их расширения как решений уравнений Эйнштейна. Подобной нежелательной модификации основных принципов можно полностью избежать в том случае, если свойство «быть решением уравнений Эйнштейна» удовлетворяет условию 3. Аналогичные замечания применимы и к вопросу о том, удовлетворяет ли условию 3 свойство «иметь замкнутые времениподобные кривые». Действительно, от статуса замкнутых времениподобных кривых по отношению к условию 3 зависит, может ли быть выполнена для построения идеальных точек (см. конец § 3) программа уточнения понятия «сингулярных точек» при помощи аппарата расширения пространства-времени.

Наконец, отметим роль наших условий в вопросе о нахождении надлежащего определения сингулярного пространства-времени. Свойство «быть геодезически полным» удовлетворяет условиям 2 и 3, но не 1. В том, что это свойство не удовлетворяет условию 1, — а подмножество полного пространства-времени не обязательно должно быть полным, — и состоит в известном смысле непригодность понятия полноты как определения несингулярного пространства-времени. С другой стороны, введя понятие расширения непосредственно в определение, мы получили в § 1 понятие несингулярности, удовлетворяющее условиям 1 и 3. [Оно не удовлетворяет условию 2; однако нельзя надеяться получить определение, удовлетворяющее условиям 1 и 2, так как «несингулярность» должна быть глобальным свойством пространства-времени.] Эта замена одних условий другими является одним из оснований для соображений, развитых в § 1.

ПРИЛОЖЕНИЕ В. ТОПОЛОГИИ, ЗАДАННЫЕ НА СОВОКУПНОСТИ ПРОСТРАНСТВ-ВРЕМЕН

Для многих рассуждений, проведенных в § 2, нам нужно было знание топологии определенных совокупностей пространств-времен. Оказывается, что существует довольно общий метод построения подобных топологий. В этом приложении мы кратко опишем этот метод и покажем, каким образом с его помощью получаются две наиболее общие топологии. Следует подчеркнуть в самом начале, что ни одна из топологий, которые будут рассмотрены здесь, не является вполне пригодной для ОТО. Тем не менее полезно иметь представление о преимуществах и недостатках различных имеющихся топологий.

Пусть M — некоторое фиксированное (связное, хаусдорфово) 4-мерное многообразие, и пусть $\mathscr{D}(M)$ — совокупность всех симметричных ковариантных тензоров класса C^∞ второго ранга, заданных на M [1]). На этой стадии нет необходимости ограничивать рассмотрение метриками сигнатуры $(+, -, -, -)$, так как множество лоренцевых метрик является подмножеством множества $\mathscr{D}(M)$ и, таким образом, будет наследовать топологию множества $\mathscr{D}(M)$. Действительно, удобнее рассматривать общий случай, так как при этом основные идеи могут быть непосредственно применены к другим типам геометрических объектов на M.

Пусть h_{ab} — произвольная положительно определенная метрика на M (с оператором ковариантного дифференцирования ∇_a), C — произвольное (не пустое) замкнутое подмножество многообразия M и, наконец, p — любое неотрицательное целое число (возможно, бесконечно большое). При помощи этих данных определим следующим образом функцию расстояния, заданную на парах элементов множества $\mathscr{D}(M)$:

$$\rho(g_{ab}, g'_{ab}) = \sup_C \sum_{n=0}^{p} 2^{-n} \frac{|g - g'|_n}{1 + |g - g'|_n}, \tag{B.1}$$

[1]) Эти рассуждения легко обобщить на метрики, относящиеся к более слабому классу дифференцируемости, но это обобщение не дало бы ничего нового и только затемнило бы суть проблемы.

где

$$|g-g'|_n=\{[\nabla_{a_1}\ldots\nabla_{a_n}(g_{rs}-g'_{rs})]\times$$
$$\times[\nabla_{b_1}\ldots\nabla_{b_n}(g_{uv}-g'_{uv})]\,h^{a_1b_1}\ldots h^{sv}\}^{1/2}.$$

Таким образом, две метрики g_{ab}, $g'_{ab}\in\mathcal{D}(M)$ являются «близкими», если их значения и значения их первых p производных близки (по отношению к метрике h_{ab}) на множестве C. [Сложная форма выражения (В.1) необходима по двум причинам: 1) чтобы обеспечить существование по крайней мере верхней границы расстояния, даже если некоторые значения $|g-g'|_n$ могут быть неограниченными на множестве C, и 2) чтобы случай $p=\infty$ не требовал специального рассмотрения. Реальная топология, получаемая на основе функции ρ, не зависит от деталей структуры выражения (В.1).]

Фиксируем теперь целое число p. Тогда каждый произвольный выбор метрики h_{ab} и множества C определит посредством функции расстояния (В.1) некоторую топологию на множестве $\mathcal{D}(M)$. Таким образом, окрестность метрики $g_{ab}\in\mathcal{D}(M)$ состоит из всех тех метрик $g'_{ab}\in\mathcal{D}(M)$, для которых $\rho(g_{ab}, g'_{ab})<\varepsilon$, где $\varepsilon>0$. Однако нас интересуют топологии, не зависящие от подобного произвольного выбора. Для получения «инвариантных» топологий поступим согласно следующему плану. Выделим некоторым инвариантным образом совокупность пар (h_{ab}, C), где каждая метрика h_{ab} положительно определена и каждое множество C замкнуто. Каждая такая пара определяет функцию расстояния (В.1), а следовательно, некоторое семейство открытых подмножеств множества $\mathcal{D}(M)$. Совокупность всех конечных пересечений и произвольных объединений всех открытых множеств, получаемых при помощи всех пар (h_{ab}, C) из нашего выделенного множества, определяет топологию на множестве $\mathcal{D}(M)$. Отсюда для фиксированного значения p проблема построения определенных топологий сводится к проблеме выбора подходящего множества пар (h_{ab}, C).

C^p-топология на множестве $\mathcal{D}(M)$ определяется совокупностью всех тех пар (h_{ab}, C), для которых множество C компактно[1]). Окрестность метрики $g_{ab}\in\mathcal{D}(M)$ состоит из всех тех метрик, для которых их значения и значения их первых p-производных близки к метрике g_{ab} в некотором компактном множестве и поведение которых вне этого компактного множества

[1]) Фиксируем данную положительно определенную метрику $\bar{h}_{ab}$. Тогда C^p-топология тождественна топологии, получаемой при помощи всех тех пар $(\bar{h}_{ab}, C)$, для которых множество C компактно.

не ограничено. C^p-топология достаточно груба, т. е. в ней существует очень мало открытых множеств. Например, в C^p-топологии подмножество множества $\mathscr{D}(M)$, состоящее из всех метрик лоренцевой сигнатуры, даже не открыто (если многообразие M некомпактно). Если $g_{ab} \in \mathscr{D}(M)$, то однопараметрическое семейство метрик λg_{ab} определяет непрерывную кривую в множестве $\mathscr{D}(M)$. Таким образом, множество $\mathscr{D}(M)$ оказывается связным. [Однако лоренцевы метрики в общем случае не образуют связное подмножество множества $\mathscr{D}(M)$. При этом обычно существует только несколько компонент, и это обстоятельство тесно связано с понятием «гомотопических классов лоренцевых метрик».] Последовательность метрик вида $\operatorname{diag}(\mu_m, -1, -1, -1)$, где $\mu_m = 1 + m_p[1 + (x - m)^2]^{-1}$ $(m = 1, 2, \ldots)$ приближает пространство Минковского в C^p-топологии. («Шишка» в метриках становится больше, по мере того как она удаляется на бесконечность.)

F^p-топология [18] на множестве $\mathscr{D}(M)$ определяется совокупностью всех тех пар (h_{ab}, C), для которых $C = M$ [1]). Таким образом, окрестность метрики $g_{ab} \in \mathscr{D}(M)$ состоит из тех метрик, для которых отклонения в каждой точке многообразия M их значений и значений их первых p производных от соответствующих значений для метрики g_{ab} находятся в определенном диапазоне, причем этот диапазон меняется на многообразии M непрерывным, но в других отношениях произвольным образом. F^p-топология достаточно тонка, а именно в ней существует очень много открытых множеств. Например, последовательность метрик вида $\operatorname{diag}(\nu_m, -1, -1, -1)$, где $\nu_m = 1 + (m^2 + x^2 + y^2 + z^2)^{-1}$ $(m = 1, 2, \ldots)$, не приближает пространства Минковского в F^p-топологии. («Шишка» в метриках остается помещенной в начале координат и уменьшается по амплитуде.)

Однопараметрическое семейство метрик λg_{ab} не непрерывно в F^p-топологии (при условии, что метрика g_{ab} имеет некомпактный носитель, например если эта метрика имеет лоренцеву сигнатуру). Наконец (предполагая опять многообразие M некомпактным), множество $\mathscr{D}(M)$ имеет в F^p-топологии много различных связных компонент. Например, множество всех тех метрик g_{ab}, для которых многообразие M имеет конечный объем, одновременно открыто и замкнуто.

В качестве частного примера опишем аналогичные топологии для функций на вещественной прямой. Каждая функция может быть представлена своим графиком. На фиг. B1

---

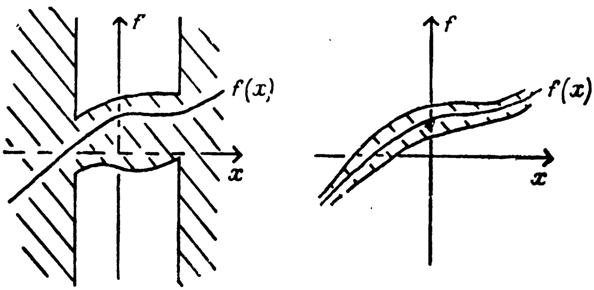
[1]) F^p-топология тождественна топологии, получаемой из *всех* пар (h_{ab}, C).

приведены графики, соответствующие функциям, находящимся в окрестности заданной функции $f(x)$.

C^{p+1}-топология более тонка, чем C^p-топология, F^{p+1}-топология более тонка, чем F^p-топология, а последняя в свою очередь более тонка, чем C^p-топология. Таким образом, имеем

$$\begin{array}{ccccccccc} C^0 & \to & C^1 & \to & C^2 & \to & \dots & \to & C^\infty \\ \downarrow & & \downarrow & & \downarrow & & & & \downarrow \\ F^0 & \to & F^1 & \to & F^2 & \to & \dots & \to & F^\infty \end{array}$$

где стрелки обозначают переход от менее тонкой топологии к более тонкой. Для некомпактного многообразия M ни одна



Фиг. B1. Окрестности функции $f(x)$ в C^0- и F^0-топологиях.
Типичная окрестность функции $f(x)$ состоит из всех тех функций, графики которых лежат в заштрихованной области.

из изображенных на диаграмме топологий не эквивалентна другой, в то время как для компактного многообразия топологии C^p и F^p совпадают. (Возникающая в результате топология вполне «приемлема» для компактного многообразия M.)

В заключение приведем несколько технических замечаний относительно топологий. Обозначим через $\mathscr{T}$ произведение множества всех положительных определенных метрик на многообразии M и множества всех непустых замкнутых подмножеств многообразия M. Тогда, как мы видели, каждое подмножество множества $\mathscr{T}$ определяет некоторую топологию на множестве $\mathscr{D}(M)$. Группа диффеоморфизмов многообразия M действует очевидным образом как группа преобразований множества $\mathscr{T}$. Подмножество множества $\mathscr{T}$, инвариантное относительно преобразований из этой группы, определяет то, что мы назвали инвариантной топологией. (В более общем виде группа диффеоморфизмов действует как группа преобразований на множестве $\mathscr{D}(M)$, а *инвариантная топология* — это топология, для которой эти преобразования являются гомеоморфизмами.) Топология хаусдорфова тогда и только тогда, когда объединение множеств C всюду плотно в многообразии M. В частности, каждая инвариантная топология (за исключением дискретной) хаусдорфова.

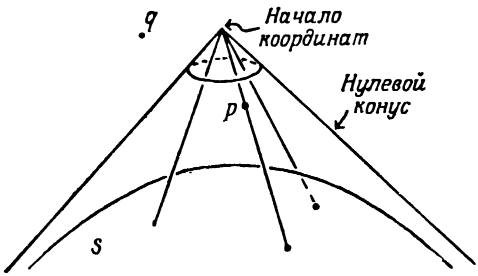
Наиболее грубая топология, которая может быть построена подобным образом, — это та топология, в которой каждое замкнутое множество C ограничено. Эта топология строго грубее, чем C^p. Наиболее тонкой топологией является F^p-топология. Вполне возможно построить инвариантные топологии, промежуточные между C^p и F^p. Например, можно рассмотреть совокупность всех таких пар (h_{ab}, C), для которых замыкание внутренности каждого множества C компактно. Наконец, несомненно существуют инвариантные топологии на множестве $\mathscr{D}(M)$, которые не могут быть построены рассмотренным образом при помощи пар (h_{ab}, C). Например, в выражении (В.1) для C^p-топологии можно добавить член $V(1+V)^{-1}$, где V — объем многообразия M относительно метрики $(g_{ab} - g'_{ab})$.

ПРИЛОЖЕНИЕ Г. ТЕОРЕМЫ О СУЩЕСТВОВАНИИ СИНГУЛЯРНОСТЕЙ

В различных теоремах о существовании сингулярностей часто используются одни и те же аргументы. Поэтому для получения общего представления о том, как проводятся доказательства и по каким причинам в теоремах возникают определенные дополнительные условия, достаточно проследить за доказательством одной подобной теоремы. Мы дадим упрощенное доказательство результата, принадлежащего Хокингу [15, 17]. Чтобы не затемнить основную идею рассуждения, некоторые технические детали будут заменены качественными аргументами. Другие рассмотрения этих теорем см., например, в работах [32, 39].

По существу все доказательство состоит из двух частей: 1) доказательства того, что при определенных условиях не существует времениподобной кривой наибольшей длины, идущей от некоторой точки до некоторой пространственноподобной поверхности, и 2) доказательства того, что это отсутствие кривой наибольшей длины приводит к геодезической неполноте.

Пусть M — пространство-время с метрикой g_{ab}, а S — произвольная пространственноподобная поверхность в M. Задав в будущем поверхности S точку p, будем выяснять, существует ли среди всех времениподобных кривых, идущих от p к S, кривая максимальной длины. (В случае индефинитной метрики времениподобная геодезическая имеет тенденцию быть наиболее длинной кривой, соединяющей точки. Кривые, которые «изгибаются» больше, чем геодезические, находятся ближе к световому конусу и поэтому имеют меньшую длину.) Чтобы построить пример, иллюстрирующий наши рассуждения, рассмотрим гиперболоид S, образованный точками, отстоящими на единичную длину от начала координат пространства Минковского



Фиг. Г1. Гиперболоид S в пространстве Минковского (одно пространственное измерение опущено).

Через любую точку p между нулевым конусом прошлого начала координат и поверхностью S проходит времениподобная кривая максимальной длины (прямые линии на фигуре). Через любую точку q, находящуюся вне этого нулевого конуса, может проходить к поверхности S времени подобная кривая произвольно большой длины.

(фиг. Г1). Как мы вскоре увидим, нулевой конус прошлого начала координат является границей между теми точками, которые могут быть соединены с поверхностью S кривой максимальной длины, и теми точками, которые не удовлетворяют этому условию.

Возвращаясь теперь к общему случаю, допустим на некоторое время, что мы получили некоторую времениподобную кривую γ, соединяющую точку p с поверхностью S и имеющую максимальную длину. Тогда кривая γ должна быть геодезической (в противном случае мы могли бы выпрямить эту кривую и тем самым увеличить ее длину) и пересекать поверхность S ортогонально ей (в противном случае мы могли бы увеличить длину кривой γ, немного меняя ее точку пересечения с поверхностью S). Таким образом, мы пришли к необходимости изучить геодезические, нормальные к поверхности S.

Проведем из каждой точки поверхности S (единственную) времениподобную геодезическую, начинающуюся ортогонально поверхности S (например, прямые линии на фиг. Г.1). Поле ξ^a единичных векторов, касательных этой конгруэнции геодезических, регулярно в некоторой окрестности поверхности S (а именно до тех пор, пока геодезические не начинают пересекаться в каустиках). Имеет место важное, лежащее в основе всех теорем существования уравнение [33], которое управляет поведением подобной конгруэнции геодезических. Определим операцию *конвергенции* (сходимости) конгруэнции геодезических: $c = -\nabla_a \xi^a$. Тогда имеем

$$\begin{aligned} \xi^b \nabla_b c &= -\xi^b \nabla_b \nabla_a \xi^a = \\ &= -\nabla_a (\xi^b \nabla_b \xi^a) + (\nabla_a \xi_b)(\nabla^b \xi^a) - R_{ba\ d}^{\ \ a} \xi^b \xi^d = \\ &= (\nabla_a \xi_b)(\nabla^b \xi^a) + R_{ab} \xi^a \xi^b, \end{aligned} \quad (\text{Г}.1)$$

где использован тот факт, что поле векторов ξ^a геодезично. Затем преобразуем два члена в правой части этого уравнения. Заметим, что в первом члене по построению

$$\nabla_a \xi_b = \nabla_{(a} \xi_{b)} \quad \text{и} \quad \xi^a \nabla_a \xi_b = 0,$$

т. е. выражение $\nabla_a \xi_b$ может рассматриваться как симметричный тензор в 3-пространстве, ортогональном вектору ξ^a. Поэтому справедливо соотношение

$$(\nabla_a \xi_b)(\nabla^b \xi^a) \geqslant \frac{1}{3} c^2,$$

представляющее собой хорошо известное матричное неравенство, а именно для любой симметричной 3×3-матрицы A имеем $\operatorname{Sp} A^2 \geqslant {}^1/_3 (\operatorname{Sp} A)^2$. Теперь предположим, что второй член в правой части (Г.1) неотрицателен. Выраженное при помощи уравнений Эйнштейна через тензор энергии-импульса материи, это допущение означает, что

$$\left(T_{ab} - \frac{1}{2} T g_{ab}\right) t^a t^b \geqslant 0$$

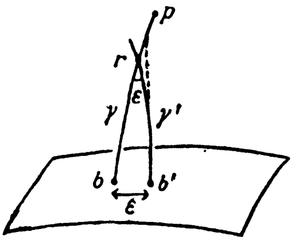
для всех времениподобных векторов t^a, т. е. известное энергетическое условие. (Например, для идеальной жидкости это условие означает, что $\rho + p \geqslant 0$, $\rho + 3p \geqslant 0$.) Уравнение (Г.1) теперь принимает окончательную форму

$$\xi^b \nabla_b c \geqslant \frac{1}{3} c^2. \quad (\text{Г}.2)$$

Уравнение (Г.2) имеет простую физическую интерпретацию: если энергетическое условие выполняется, то имеет место необратимое увеличение c вдоль конгруэнции, т. е. времениподобные геодезические должны сходиться.

Основная идея заключается в применении уравнения (Г.2), чтобы проверить катастрофическую тенденцию для нормальных геодезических, исходящих из поверхности S. Один из путей состоит в использовании предположения (как в случае, показанном на фиг. Г1), что геодезические сходятся уже тогда, когда они покидают поверхность S, т. е. мы теперь предполагаем, что $c \geqslant c_0 > 0$ на поверхности S, где c_0 — константа. Тогда из уравнения (Г.2) вытекает, что значение c становится бесконечным не далее как на расстоянии $s_0 = 3/c_0$ от поверхности S вдоль каждой (направленной в будущее) нормальной геодезической. Бесконечное значение c служит сигналом того, что наши геодезические уже начали пересекаться. Тот факт, что каждая геодезическая доходит до каустики на расстоянии не более s_0 от поверхности S, еще не означает сингулярного поведения для части самого пространства-времени, так как каустики могут образовываться и в регулярной области пространства-времени (например, в начале координат на фиг. Г1).

Причина, по которой нас интересуют каустики, несколько более тонкая. Пусть γ — одна из времениподобных геодезических, нормальных к поверхности S; точку p выберем за значением $s = s_0$ на кривой γ (фиг. Г2). Эта кривая γ проходит через кау-



Фиг. Г2. Геодезическая γ' пересекает геодезическую γ до того, как γ достигает точки p.

Разность длин $prb - prb'$ порядка ε^3; поэтому путь из точки p до поверхности S по штриховой линии (оканчивающейся в точке b') имеет большую длину, чем путь prb вдоль кривой γ.

стику до того, как она достигнет точки p, и поэтому найдется некоторая соседняя геодезическая γ' (которая также начинается нормально к поверхности S), пересекающая кривую γ в точке r ранее точки p. Угол между этими двумя геодезическими в точке r порядка ε, в то время как разность расстояний вдоль обеих геодезических от точки r до поверхности S порядка ε^3 [1]). Отсюда теперь следует, что, «спрямляя» угол (пунктирная линия на фиг. Г2) ломаной геодезической prb', мы получим времениподобную кривую, идущую от точки p до поверхности S и имеющую длину, большую, чем геодезическая γ.

Подводя итоги сказанному, заключаем, что: 1) если существует времениподобная кривая наибольшей длины, идущая от точки p до поверхности S, то это должна быть нормальная геодезическая, и 2) если какая-либо нормальная геодезическая имеет длину большую, чем $3/c_0$, то она не является времениподо-

[1]) Разложим разность расстояний по степеням ε. Член порядка ε исчезает, так как геодезическая γ пересекает поверхность S нормально. Член порядка ε^2 исчезает, поскольку кривизна поверхности S в точке b как раз такова, что геодезическая γ' пересекается с геодезической γ в точке r. Поэтому первый неисчезающий член имеет порядок ε^3.

добной кривой наибольшей длины, идущей от точки p до поверхности S. Другими словами, мы показали, что *если выбрать любую точку p на расстоянии от поверхности S, большем, чем* $3/c_0$ (вдоль любой времениподобной кривой), *то не существует времениподобной кривой максимальной длины, идущей от точки p до поверхности S.* (Как раз эта ситуация изображена на фиг. Г1.)

Чтобы эту трудность нахождения кривых максимальной длины, идущих к поверхности S, обратить в противоречие, сделаем теперь заключительное предположение о том, что поверхность S является поверхностью Коши [11, 15, 30], т. е. что каждая времениподобная кривая (без конечных точек) пересекает поверхность S точно один раз. Гиперболоид S, показанный на фиг. Г1, не представляет собой поверхности Коши (некоторые времениподобные кривые, например исходящие из точки q, не достигают поверхности S), в то время как времениподобная плоскость в пространстве Минковского является поверхностью Коши. Тот факт, что поверхность S представляет собой поверхность Коши для пространства M, интуитивно означает, что полная эволюция пространства-времени M полностью и единственным образом определяется начальными данными на поверхности S. Пусть p — произвольная точка, принадлежащая будущему поверхности S. Поскольку *каждая* времениподобная кривая исходящая из точки p, пересекает поверхность S, должна существовать кривая максимальной длины, идущая от точки p до поверхности S. (Это утверждение, которое интуитивно представляется разумным, доказано в работах [11, 15].) Но выделенное выше курсивом утверждение говорит о том, каким образом можно найти точку p, через которую нельзя провести к поверхности S кривую максимальной длины. Разрешить это противоречие можно лишь одним способом — сделать вывод о том, что подобную точку p найти просто невозможно, т. е. что ни одна направленная в будущее времениподобная кривая не имеет длины, большей чем $3/c_0$. Таким образом, доказана

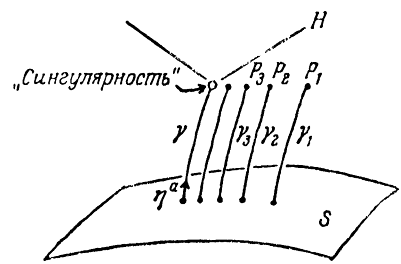
Теорема Г1 (Хокинга). Пусть M — пространство-время, тензор Риччи которого удовлетворяет энергетическому условию $R_{ab}t^a t^b \geqslant 0$ для всех времениподобных векторов t^a. Пусть также M содержит поверхность Коши S, конвергенция которой c удовлетворяет условию $c \geqslant c_0 > 0$. Тогда каждая направленная в будущее времениподобная кривая, исходящая из поверхности S, имеет длину, не большую чем $3/c_0$. (В частности, *каждая* времениподобная геодезическая в M неполна.)

Теорема Г1 достаточно слаба. Предположим, что мы смогли найти в пространстве-времени M пространственноподобную поверхность S, конвергенция которой везде больше некоторого значения c_0 (например, поверхность, показанная на фиг. Г1). По мере того как пространство-время эволюционирует от поверхности S, мы будем наблюдать за появлением сингулярностей, однако нас вполне может постигнуть разочарование: после определенной эволюции пространства-времени может в конце концов оказаться, что поверхность S не являлась поверхностью Коши. (Например, гиперболоид, показанный на фиг. Г1, не является поверхностью Коши для пространства Минковского.) Проблема заключается в том, что, изучая только окрестность поверхности S, невозможно определить, окажется ли в конечном счете поверхность S поверхностью Коши или нет. Таким образом, теоремы о существовании сингулярностей, использующие поверхность Коши, вообще значительно слабее тех теорем, для которых наличие поверхности Коши не является необходимым. Теперь мы коротко укажем, каким образом теорема Г1 может быть видоизменена таким образом, чтобы исключить предположение о существовании поверхности Коши.

Если поверхность S не является поверхностью Коши, то существует некоторая область, называемая областью зависимости от поверхности S (или областью эволюции Коши поверхности S) [6, 11, 15, 32], которая состоит из тех точек, исходя из которых все направленные в прошлое времениподобные кривые наверняка пересекают поверхность S. (На фиг. Г1 областью зависимости от поверхности S является область, расположенная между поверхностью S и нулевым конусом прошлого начала координат.) Наши предыдущие аргументы справедливы в пределах этой области зависимости. Граница будущего области зависимости носит название горизонта Коши [11, 17, 32]. Это всегда нулевая поверхность (возможно, с «углами»: горизонтом Коши поверхности S, показанной на фиг. Г1, является нулевой конус прошлого начала координат). Основная идея заключается в том, чтобы рассмотреть, что́ же происходит на этом горизонте Коши. Допустим, что поверхность Коши компактна (т. е. «конечна»: сфера и тор компактны, плоскость и цилиндр нет). Главный вопрос состоит в следующем: является ли горизонт Коши H поверхности S также компактным? Любой ответ приводит к трудностям.

Предположим сначала, что горизонт Коши H компактен. Но H — нулевая поверхность. Нулевая кривая, проведенная на компактной поверхности H, будет извиваться по поверхности до тех пор, пока она в конце концов не вернется обратно и не подойдет бесконечно близко к себе самой. Эта ситуация представляет собой один из типов нарушения причинности (почти замкнутые нулевые кривые), практически настолько же серьезный, как существование замкнутых времениподобных кривых. (Эта первая возможность реализуется в пространстве Тауба — НУТ [28].)

Предположим теперь, что горизонт Коши H некомпактен. Тогда можно найти последовательность точек p_1, p_2, p_3, ... (фиг. Г3), не имеющую точки сгущения. Для каждой точки p_i существует времениподобная геодезическая γ_i, длина которой от точки p_i до поверхности S максимальна и (поэтому) меньше, чем s_0. Каждая из этих геодезических единственным образом определяется своим единичным касательным вектором, заданным в точке пересечения кривой с поверхностью S. Таким образом, имеется последовательность векторов, нормальных S.



Фиг. Г3. Если поверхность H некомпактна, то выберем между поверхностями H и S последовательность точек p_i, не имеющую точки сгущения; предельная кривая γ соответствующей последовательности максимальных геодезических, соединяющих точку p с поверхностью S, определяет неполную геодезическую

Но поверхность S по предположению компактна, поэтому эта последовательность сходится к некоторому единичному вектору η^a, нормальному S. Если бы геодезическая, соответствующая этому вектору, могла быть продолжена до длины, большей чем s_0, то можно было бы построить точку сгущения для последовательности точек p_i (фиг. Г3). Отсюда следует, что эта геодезическая γ должна быть неполной. Таким образом, доказана

Теорема Г2 (Хокинга). Пусть M — пространство-время, тензор Риччи которого удовлетворяет энергетическому условию, и пусть M содержит компактную пространственноподобную поверхность S, конвергенция которой $c \geqslant c_0 > 0$. Тогда, если предположить, что M не содержит почти замкнутых нулевых кривых, то M должно содержать некоторую неполную времениподобную геодезическую.

Отметим, что теорема Г2 отличается от теоремы Г1 тремя существенными моментами: 1) не предполагается наличие поверхности Коши, 2) необходимо предположить компактность поверхности S и 3) доказывается существование только одной неполной геодезической, а не неполнота всех геодезических. На самом деле теорема Г2 справедлива даже в том случае, если не исключать существования почти замкнутых нулевых кривых [17].

ПРИЛОЖЕНИЕ Д. ЗАДАЧИ

Приведенные ниже задачи в большинстве своем достаточно трудны. Насколько мне известно, все они еще не решены. (Желающие решать более легкие задачи могут обратиться к многочисленным утверждениям в тексте этого обзора, особенно в приложениях, которые приводились без доказательств.) Большинство проблем, затронутых в задачах, уже упоминалось в тексте, но мы повторим их здесь, чтобы дать полную картину некоторых направлений наступления на общую проблему сингулярностей в ОТО. Отметим, что суть многих задач заключается просто в нахождении достаточно точных определений некоторых интуитивных понятий.

Определение сингулярного пространства-времени

1. Исследуйте определения сингулярного пространства-времени, основанные на учете «приливных» сил, испытываемых определенным образом выбранными наблюдателями. Справедливо ли утверждение, что если приливные силы по предположению не возрастают слишком сильно, то некоторые типы неполноты (см., например, приложение А) становятся эквивалентными?

2. Найдите контрпримеры с целью показать, что импликации в (А. 1) необратимы. (Почти наверняка можно видоизменить пример, приведенный в работе [8], и показать, что из G-полноты не вытекает T_n-полнота при любом n.)

3. Найдите пример пространства-времени, которое нуль-неполно, но пространственноподобно и времениподобно полно. (Приведенное утверждение представляет собой единственный случай этого типа, для которого не известно контрпримера; см. [8, 23].)

4. Исследуйте определения сингулярного пространства-времени, основанные на понятии неполноты Мизнера [26, 35].

5. Найдите определение сингулярного пространства-времени, интуитивно удовлетворительное, достаточно простое, а также такое, чтобы существующие теоремы о сингулярностях могли бы быть интерпретированы исходя из этого нового определения.

Существование сингулярностей

6. Докажите следующее утверждение (или найдите контрпример). Каждое нерасширяемое пространство-время, содержащее компактную пространственноподобную 3-поверхность, тензор энергии-импульса которого удовлетворяет надлежащему энергетическому условию, либо G-сингулярно, либо плоско. (Энергетическое условие должно быть достаточно сильным, чтобы исключить модель вселенной Эйнштейна. Возможно, сначала следует доказать, что 3-поверхность не может быть в топологическом смысле 3-сферой.)

7. Докажите, что при $\Lambda < 0$ не существует ни одной несингулярной замкнутой модели вселенной.

8. Докажите с использованием надлежащей топологии, что на фиг. 2 ни при каких обстоятельствах не может быть серых областей. (Подобный результат, по всей вероятности, можно доказать без использования каких-либо энергетических условий или других теорем, опираясь только на определение сингулярного пространства-времени.)

9. Найдите приемлемую топологию на множестве $\tilde{\mathscr{D}}(M)$. Докажите в этой топологии некоторые теоремы, отражающие наше интуитивное рассмотрение понятий замкнутых и открытых моделей вселенной. (В случае открытой модели, возможно, удастся доказать некоторые теоремы об асимптотически плоском пространстве-времени [31].)

10. Возможно, что попытки интерпретировать фиг. 2 в топологических терминах необоснованны. Обдумайте какую-либо другую структуру на множестве $\tilde{\mathscr{D}}(M)$, для которой можно было бы сделать точные утверждения. (Например, можно попытаться определить меру на множестве $\tilde{\mathscr{D}}_c(M)$ и доказать, что множество несингулярных метрик имеет меру нуль.)

11. Охарактеризуйте некоторым простым образом все инвариантные топологии на множестве $\mathscr{D}(M)$, которые можно построить на основе нашего метода, использующего пары (h_{ab}, C). Или, в более общем виде, найдите *все* инвариантные топологии на множестве $\mathscr{D}(M)$. Применимы ли какие-либо из этих топологий для обсуждения фиг. 2?

12. Трудность с F^p-топологией заключается в том, что, поскольку метрика h_{ab} на бесконечности может стать очень малой, мы склонны требовать, чтобы различные метрики в окрестности метрики g_{ab} были «слишком близки» к метрике g_{ab} вблизи бесконечности. Это замечание предполагает, что мы рассматриваем такие пары (h_{ab}, C), для которых $C = M$, а метрика h_{ab}

полна. Можно ли подобным образом получить топологию, отличную от F^p-топологии? Будет ли она представлять интерес?

13. Найдите точный способ формулировки следующего утверждения: «Положительная космологическая константа имеет тенденцию создавать меньше сингулярных решений». Докажите это утверждение.

Свойства сингулярностей

14. Найдите способ определения сингулярных точек, который мог бы соединить простоту конструкции, использующей понятие идеальных точек, с богатством структуры g-границы. (Возможно, наиболее вероятен успех при использовании определения, непосредственно связанного с энергетическим условием.) Разделите полученные подобным образом сингулярные точки на небольшое количество типов.

15. В конце § 3 мы наметили программу уточнения существующих представлений о сингулярных точках. Реализуйте эту программу для понятия g-границы и (или) для понятия идеальных точек.

16. Найдите определение сингулярных точек, основанное на теореме Кларка о погружении [5].

17. До какой степени функции ускорения, характеризующие сингулярное пространство-время, к нему приспособлены? Можно ли классифицировать сингулярности согласно функциям ускорения, приспособленным или не приспособленным к данному пространству-времени?

Разные задачи

18. Во множестве $\tilde{\mathscr{D}}(M)$ физически тождественные метрики изображаются многими различными точками. Изучите множество физически различных метрик. Например, обладает ли это множество какими-либо интересными топологиями?

19. Вид определения «замкнутой модели вселенной», приведенного в § 2, обусловлен тем обстоятельством, что в некоторых теоремах о сингулярностях используется существование компактной пространственноподобной 3-поверхности. Найдите более подходящее определение, не изменяющее, однако, качественных особенностей фиг. 2. (В частности, было бы желательно, чтобы компактные модели вселенной были «замкнутыми».)

20. Определите, удовлетворяют ли свойства «быть свободным от источников решением уравнений Эйнштейна» и «не иметь замкнутых времениподобных кривых» условию 3 приложения Б. (Аналогично для свойств «удовлетворять энергетическому условию», «иметь вектор Киллинга» и многих других свойств пространства-времени.)

ЛИТЕРАТУРА

1. *Avez A.*, Ann. Inst. Fourier, **13**, 105 (1963).
2. *Bertotti B.*, Phys. Rev., **116**, 1331 (1959).
3. *Boyer R. H., Lindquist R. W.*, Journ. Math. Phys., **8**, 265 (1967).
4. *Choquet-Bruhat Y., Geroch R.*, Comm. Math. Phys. (в печати).
5. *Clarke C. J. S.*, On the Global Isometric Embedding of Pseudo-Riemannian Manifolds, preprint Camb. Univ., 1969.
6. *Courant R., Hilbert D.*, Methods of Mathematical Physics, vol. II, New York, 1965 (см. перевод первого издания: *Р. Курант, Д. Гильберт*, Методы математической физики, том. 2, М — Л., 1951).
7. *Ehresmann C.*, Colloque de Topologie Algebrique (espaces fibres), Bruxelles, 1951.
8. *Geroch R.*, Ann. Phys. **48**, 526 (1968).
9. *Geroch R.*, Journ. Math. Phys., **9**, 450 (1968).
10. *Geroch R.*, в книге Battelle Recontres (eds. C. Dewitt, J. Wheeler), New York, 1968.
11. *Geroch R.*, Journ. Math. Phys. (в печати).
12. *Geroch R.*, Comm. Math. Phys. (в печати).
13. *Geroch R., Kronheimer E. H., Penrose R.*, Proc. Roy. Soc. (в печати).
14. *Graves J. C., Brill D. R.*, Phys. Rev., **120**, 1507 (1960).
15. *Hawking S. W.*, Proc. Roy. Soc., **294A**, 511 (1966).
16. *Hawking S. W.*, Proc. Roy. Soc., **295A**, 490 (1966).
17. *Hawking S. W.*, Proc. Roy. Soc., **300A**, 187 (1967).
18. *Hawking S. W.*, Proc. Roy. Soc., **308A**, 433 (1969).
19. *Hawking S. W., Penrose R.*, Proc. Roy. Soc. (в печати).
20. *Jordan P., Ehlers J., Kundt W.*, Akad. Wiss. der Lit. Mainz (1960).
21. *Kelly J. L.*, General Topology, New York, 1955 (см. перевод: *Дж. Келли*, Общая топология, М., 1968).
22. *Kruskal M. D.*, Phys. Rev., **119**, 1743 (1960).
23. *Kundt W.*, Zs. Phys., **172**, 488 (1963).
24. *Maitra S. C.*, A Stationary Dust-Filled Cosmological Solution with $\Lambda = 0$ and Without Closed Timelike Lines, preprint Univ. of Maryland, 1965.
25. *Melvin M. A.*, Phys. Letters, **8**, 65 (1964).
26. *Misner C. W.*, Journ. Math. Phys., **4**, 924 (1963).
27. *Misner C. W.*, The Mix-Master Universe, preprint Univ. of Maryland, 1969.
28. *Newman E. T., Tamburino L., Unti T.*, Journ. Math. Phys., **4**, 915 (1963).
29. *Nowacki V. W.*, Comm. Math. Helv., **7**, 81 (1934).
30. *Penrose R.*, Phys. Rev. Letters, **14**, 57 (1965).
31. *Penrose R.*, Proc. Roy. Soc., **284A**, 159 (1965).
32. *Penrose R.*, в книге Battelle Recontres (eds. C. Dewitt, J. Wheeler), New York, 1968.
33. *Raychaudhuri A.*, Phys. Rev., **98**, 1123 (1955).
34. *Robinson I.*, Bull. Acad. Polon. Sci., **7**, 351 (1959).
35. *Shepley L.*, SO(3, R) Homogeneous Cosmologies, PhD thesis, Dept of Phys., Princeton, 1965.
36. *Spanier E. H.*, Algebraic Topology, New York, 1966 (см. перевод: *Е. Спеньер*, Алгебраическая топология, М., 1972).
37. *Synge J. L.*, Relativity: The General Theory, Amsterdam, 1960 (см. перевод: *Дж. Синг*, Общая теория относительности, М., 1963).
38. *Taub A. H.*, Ann. Math., **53**, 472 (1951).
39. *Thorne K. S.*, Gravitational Collapse, a Review-Tutorial Article, 1968.

Статья 2

Квантование общей теории относительности

Д. БРИЛЛ, Р. ГОУДИ

D. R. Brill, R. H. Gowdy, Rep. Progr. Phys., **33**, 413 (1970).

Сформулировано несколько типов квантовых ОТО. Каждая из них содержит элементы произвольности и неоднозначности, а также технические трудности, что делает их неудовлетворительными. Однако структура этих теорий дала много сведений об ОТО как динамической системе и стимулировала развитие новых подходов к квантовой теории.

С точки зрения общей символики пересмотрены такие подходы к квантованию гравитационного поля в отсутствие материи, как канонический формализм, суммирование по историям, теория источников. Проводится обсуждение квантовой теории и ОТО с целью сделать обзор достаточно полным и доступным для читателей с общефизической подготовкой. Квантование открытых пространственно-временных геометрий (гравитонное рассеяние) проводится достаточно подробно, чтобы показать основную математическую структуру формализма. При обсуждении квантования замкнутых вселенных (квантовая космология) особое внимание уделяется понятию суперпространства и построению конечномерных моделей квантовых теорий. Отдельно рассматривается также структура многообразия квантового функционала состояний в ОТО.

§ 1. ВВЕДЕНИЕ

Гравитационное поле в своем классическом описании в эйнштейновской общей теории относительности (ОТО) обнаруживает замечательные аналогии, а также важные различия с лоренц-ковариантными теориями поля остальной части физики. Такие особенности гравитационного поля, как существование независимых возбуждений, несущих положительную энергию [4, 5, 17, 18], указывают на необходимость квантования. Другие особенности, в частности связанные с существенной геометрической природой гравитационного поля, ответственны как за трудности, так и за специальный интерес к программе квантования.

Легко показать качественно типичные особенности квантованного гравитационного поля. Во многих отношениях гравитоны должны вести себя как безмассовые частицы со спином 2, очень слабо взаимодействующие с материей и друг с другом (вследствие нелинейности гравитационных уравнений поля) [37]. Из-за такого слабого взаимодействия флуктуации гравитационного вакуума должны быть значительными только на чрезвычайно малых расстояниях порядка планковской длины $(hG/c^3)^{1/2} = 10^{-33}$ см [73, 76]. Эффекты такой величины пренебрежимо малы для любого обычного физического процесса и в принципе могут быть неизмеримы [63]. Однако специальные геометрические и нелинейные особенности точного квантованного гравитационного поля вполне могут приводить к важным физическим результатам, чего нельзя ожидать от простого приближения безмассового поля спина 2.

В настоящее время не имеется удовлетворительной строгой квантовой теории гравитационного поля. Существуют лишь разрозненные куски, собранные в двух разделах — ОТО и квантовой теории поля. Отсутствует какое-либо согласие в выборе правильного пути объединения этих разделов в единую картину. Некоторые из подходов к проблеме квантовой ОТО можно охарактеризовать следующим образом.

1. Один из законченных разделов совпадает с частью другого.

2. Существующие разделы, как они есть, полностью подходят друг к другу.

3. Существующее взаимодействие частей в разделах неприемлемо (постулаты не соответствуют возникающим понятиям).

4. Разделы содержат фиктивные части, которые окажутся ненужными в полной картине.

5. Разделы, имеющиеся в настоящее время, включают некоторые ложные части.

6. Для правильного объединения разделов необходимы важные дополнительные части.

Утверждение 1 представляет собой раннюю точку зрения Эйнштейна, заключающуюся в том, что к принципу относительности не следует добавлять квантовый принцип, поскольку последний должен вытекать из теории относительности. Теперь подавляющее большинство авторов рассматривают эти два принципа на равной основе и считают, что задача состоит в том, как их соединить. Самые ранние исследования этой задачи [76] были проведены исходя из утверждения 2, но эта точка зрения в настоящее время в основном оставлена. Следующие два утверждения характеризуют сущность современных исследований. Здесь также возможны различные точки зрения на то, какие из понятий должны быть пересмотрены или отброшены. Утверждение 5 предполагает, что для того, чтобы стало возможным квантование гравитации, необходимо пересмотреть современные теории, описывающие ряд известных явлений. Именно эта точка зрения вызвала недавно возрождение интереса к линейным (но нелокальным) теориям гравитации в надежде, что они смогут лучше подойти к программе квантования [20]. Наконец, последнее утверждение выражает противоположное мнение, что квантование ОТО преждевременно и необходимо дальнейшее развитие квантовой теории поля и проведение дополнительных экспериментов по ОТО.

При обзоре современного состояния квантовой ОТО мы сталкиваемся с большим разнообразием концептуальных платформ — различными подходами к каноническому формализму, разными процедурами обоснования квантовой операторной алгебры, методами суммирования по историям, теорией источников Швингера и теорией гравитации как безмассового поля спина 2. Даже в рамках избранной платформы не всегда имеется согласие о наилучшем сопоставлении постулатов и полученных понятий. Каждая точка зрения требует значительных затрат труда для построения формализма. Таким образом, необходимо найти компромисс между крайними случаями поверхностного обсуждения всех подходов и детального рассмотрения только одного из них. Наш путь будет заключаться в кратком вступительном обзоре канонического квантования, поскольку это самый старый и все же наиболее широко используемый подход, правил фейнмановского суммирования по историям, так как они часто появляются в реальных расчетах, и теории источников Швингера, так как это относительно недавно предложенный подход. Тем не менее для детального обсуждения квантовой ОТО мы вводим формализм, который объединяет суммирование по историям и каноническое квантование. Это позволит нам разобрать больше материала и с большей глубиной, чем было бы возможно в других случаях.

В § 2 настоящего обзора показывается необходимость квантования ОТО; § 3 посвящен краткому обзору ОТО, и в нем вводятся обозначения «3 + 1», расщепляющие пространство и время. Эти обозначения надо рассматривать как одно из главных достижений, появившихся в результате попыток квантования ОТО. В § 4 дается обзор современной ситуации в канонических или гамильтоновых формулировках ОТО. Эти формулировки, которые поставили ОТО в соответствие всем остальным динамическим теориям, являются другим важным достижением, возникшим из программы квантования. В § 5 рассматривается техника квантования; основной результат этого параграфа — построение квантовой теории, если известна классическая. Этот результат в некоторой степени отличается от простого понимания квантовой теории. В этом параграфе также вводится понятие представления функционалом по историям, являющееся основным для дальнейшего и дающее единую основу для всех подходов к квантованию. В § 6 рассматривается квантование открытых вселенных — единственная область квантовой ОТО, где получаются конкретные результаты с хорошо определенными интерпретациями. В открытых вселенных основная цель обсуждения — гравитонное рассеяние и вычисление S-матрицы. § 7 связан с квантованием замкнутых вселенных. Здесь можно встретить все те ситуации, которые вызывают чрезвычайные трудности при квантовании ОТО. В этой области квантовой ОТО нет группы Пуанкаре, нет S-матрицы и нет единодушия в физической интерпретации. Это область чистого формализма и догадок. Однако один результат, по-видимому, вытекает из формализма: важность понимания структуры многообразия квантового функционала состояния. Некоторые аспекты этого многообразия (суперпространства) рассматриваются в § 8.

§ 2. НЕОБХОДИМОСТЬ КВАНТОВАНИЯ ГРАВИТАЦИИ

Прежде чем начать подробное обсуждение ряда подходов к квантовой ОТО, мы кратко опишем проблемы, которые можно решить с ее помощью, и новые точки зрения, которые могут отсюда возникнуть.

1. Неприемлемые трудности в неквантовой ОТО

В течение многих лет известные физические следствия уравнений Эйнштейна основывались на немногих специальных решениях и приближенных процедурах. Фактически нельзя было сделать никаких заключений об общем решении или классе решений ненулевой меры этих чрезвычайно нелинейных уравнений. Недавно стало возможно обрисовать одну общую характерную черту классического гравитационного поля: было найдено, что при развитии во времени большого класса классических решений существуют геометрические сингулярности, хотя многие из этих решений представляют конфигурации гравитационного поля, по-видимому встречающиеся в природе. Одним из примеров такого решения является типичная космологическая модель с замкнутыми пространственноподобными сечениями, которая, насколько мы знаем, есть точное представление нашей вселенной. Можно показать [48], используя только очень общие предположения типа положительности энергии, что независимо от специальных свойств симметрии метрика такой вселенной не может оставаться регулярной вне конечного промежутка собственного времени (гравитационный коллапс). После того как достигнуто состояние коллапса, для предсказания, что́ же произойдет дальше, уже нельзя использовать уравнения Эйнштейна. Таким образом, мы неизбежно приходим к заключению, что классическая теория неполна, так как в некоторых совсем общих ситуациях она не может предсказать временно́е поведение гравитационного поля за границами определенной стадии.

2. Основные принципы современной физики требуют квантования гравитационного поля

Трудность, связанная с гравитационным коллапсом, похожа на трудность, связанную с коллапсом классического атома Резерфорда вследствие излучения [92]. В настоящее время мы не уверены, что квантовый принцип может устранить первую трудность, как он устранил последнюю, но все же квантование является наиболее очевидным этапом, опущенным в современной картине мира, а более радикальное классическое изме́нение теории (например, теория Йордана — Бранса — Дикке) вносит проблемы гравитационного коллапса, аналогичные проблемам коллапса в чистой ОТО. Более того, не найдено ни одного последовательного пути, позволяющего связать неквантованное гравитационное поле с источником квантованной материи; *c*-числа, которые можно образовать из квантованного поля материи и которые были бы подходящими источниками классического гравитационного поля (например, среднее значение тензора энергии-импульса), либо не имеют смысла для частных экспериментов, либо не удовлетворяют закону сохранения, когда измерение редуцирует волновой пакет, в зависимости от используемой интерпретации квантовой теории.

Итак, квантовое поведение материи влечет за собой описание гравитационного поля, генерируемого материей, также посредством амплитуд вероятности. Более важный вопрос связан с квантованием степеней свободы, не зависящих от материи. Годы усилий в исследовании классических гравитационных волн привели к неизбежному выводу, что гравитационное поле действительно имеет такие степени свободы и может обмениваться энергией с материей [80]. Было бы неразумно приписать гравитационному полю некое специальное свойство, позволяющее его степеням свободы избежать квантования, которое распространяется на всю остальную физику. Более естественно потребовать, чтобы эти степени свободы также описывались амплитудами вероятности и подчинялись квантовому принципу, т. е. чтобы гравитационная энергия передавалась квантами, как и все другие виды энергии.

3. Квантование ОТО ставит вопросы измерения и интерпретации, которые могут привести к лучшему пониманию квантовой теории

Квантование ОТО не только логически необходимо, но также интересно с теоретической точки зрения. Например, не очевидно, как измерять квантованные гравитационные полевые величины. В соответствующем анализе электромагнитных полевых величин, проведенном Бором и Розенфельдом [16], существенно, что отношение заряда к массе пробных частиц не определяется теорией, но может быть выбрано произвольно малым. В противоположность этому аналогичная гравитационная величина — отношение гравитационной массы к инертной массе — определяется и фиксируется принципом эквивалентности [1]).

Другая интересная проблема связана с группой общих преобразований координат в ОТО. Физический смысл этой группы совершенно отличен от физического смысла группы Лоренца в обычных теориях поля, и ее нельзя непосредственно использовать для отбора подходящих наблюдаемых для квантования. Мы обретем новое понимание принципа общей ковариантности, когда узнаем, переносится ли он и в какой степени в квантовую теорию. С координатной инвариантностью связано появление связей как части уравнений поля. Поскольку для этих нелинейных связей нельзя удобным образом выразить зависимые степени свободы через независимые, необходимо было развить новые методы исследования. Наконец, ОТО — единственная классическая теория, которая позволяет рассматривать всю вселенную в целом. Соответствующий квантовый вектор состояния также должен описывать всю вселенную. Интересной проблемой в интерпретации квантовой теории является выяснение физического смысла такого вектора состояния, для которого не существует классических внешних измерительных приборов.

4. Квантование ОТО обещает богатые возможности

В последнее десятилетие стало общепринятым, что классическая ОТО гораздо более многогранна и содержательна, чем

[1]) Подробное рассмотрение проблемы измерений в приближении слабого поля дано в работе де Витта [21]. Двумя важными результатами этой работы являются подтверждение формализма квантовой гравитации и предельное значение 10^{-32} см для наименьшей допустимой величины измеряемой области.

обычные линейные теории поля. Из чистых гравитационного и электромагнитного полей можно построить объекты, имеющие по существу все свойства (массу, заряд, законы движения и т. д) классических тел. Поэтому естественно ожидать аналогичного богатства и от квантовой ОТО (квантовой геометродинамики). Уилер [90, 92] дал наиболее подробное и образное описание ожидаемых следствий. Здесь мы упомянем лишь небольшую их часть.

Одной из наиболее важных характерных черт ОТО является возможность (фактически даже необходимость) рассматривать пространственно-временные геометрии, отличные от геометрии Минковского. Следовательно, естественно поставить вопрос о том, что может предсказать эта теория относительно возможности существования различных топологических структур пространства-времени. На этот вопрос в классической теории был получен ответ, заключающийся в том, что существует множество топологий пространственных сечений пространства-времени, для которых имеются решения уравнений Эйнштейна (по крайней мере для конечного промежутка времени), но трехмерная пространственная топология не может меняться от одного момента времени к другому [42]. Пространства с топологией, имеющей ряд ручек или «дырок», полезны для построения классических моделей массы и заряда, но вследствие неприемлемого отношения заряда к массе они не могут служить реальными моделями элементарных частиц. Соответствующий квантовый объект должен проникать через классический «барьер», препятствующий изменению топологии, т. е. пространство резонирует между различными топологиями. Подобные возбуждения могут обладать крупномасштабными коллективными свойствами, более близкими к соответствующим свойствам элементарных частиц. Если такие возбуждения возможны, то они должны фактически содержаться в общем квантовом состоянии и, следовательно, быть элементарными частицами. Согласно тому же принципу экономии мысли, заряд должен объясняться не как таинственная негеометрическая субстанция, а как силовые линии, пойманные в ловушку изменяющейся топологией, поскольку подобный ловушечный эффект *будет* иметь место в квантовом состоянии.

Чрезвычайная слабость и малость характерного масштаба квантовых гравитационных эффектов, проявляющиеся при описании элементарных частиц в геометродинамике, преодолеваются путем допущения коллективных возбуждений (частиц), включающих в себя большое число элементарных возбуждений (горловин). На первых этапах развития такой теории можно допустить существование обоих классических полей с нулевой массой покоя — гравитацию *и* электромагнетизм — в качестве элементарных строительных блоков. Но уже имеются дальнейшие обобщения: Уилер предположил, что сам фотон может являться коллективным описанием некоторых особенностей квантованной геометрии. Если это так, то уравнения Максвелла можно получить как классический предел этих особенностей чистой геометрии. Трудно себе представить более яркое свидетельство корректности квантовой геометродинамики, чем подобный вывод.

Наконец, квантовая геометродинамика требует применения квантового принципа к наиболее «классической» из всех сущностей — к вселенной как целому. Действительно, модель вселенной по многим причинам является простейшим применением квантовой ОТО, поскольку все ее крупномасштабные особенности почти полностью обусловлены гравитацией. Слабость гравитационного взаимодействия преодолевается за счет большого числа частиц и за счет малости времени первоначального расширения или конечного сжатия, в течение которого важны квантовые эффекты. Здесь возникают серьезные космологические и даже метафизические проблемы, среди которых не последним является вопрос о том, что было до «большого взрыва», положившего начало истории вселенной, и что произойдет после конечного коллапса, — тем самым приоткрывается дверь, наглухо закрытая в классической теории вследствие неизбежной сингулярности. Расчеты, выполненные де Виттом и Уилером на основе квантовых моделей, показывают, что, строго говоря, этот вопрос лишен смысла, так как одним из следствий квантового состояния нечетко определенной геометрии является то, что само время теряет смысл, а вместе с ним понятия «до» и «после» [24, 92]. Все же, в каком-то другом смысле, вселенная, по-видимому, «отскакивает» от сингулярного состояния: типичное квантовое состояние одновременно содержит много расширяющихся и коллапсирующих компонент [24]. Выдвигались предположения, что свойства элементарных частиц и другие физические константы могут изменяться от одной компоненты к другой [94]. Так, свойства наблюдаемой нами вселенной частично могут определяться требованием, чтобы в ней могла развиваться жизнь. Совершенно очевидно, что ответы на эти вопросы, которые может дать квантовая геометродинамика, имеют важные следствия, выходящие за пределы ОТО и даже физики в целом.

§ 3. КРАТКОЕ ИЗЛОЖЕНИЕ ТЕОРИИ ЭЙНШТЕЙНА

1. Геометрические предпосылки

Общая теория относительности — это теория, доводящая до своего логического завершения фундаментальное следствие принципа эквивалентности, утверждающее, что теория гравитации должна в то же время быть теорией геометрии пространства-времени. Уравнения Эйнштейна ОТО определяют, в частности, возможные пространственно-временные геометрии, однако очень многое об этих геоометриях предполагается с самого начала. Классическая теория предполагает, что физические события можно сопоставить с континуумом точек, прилегающих достаточно гладко друг к другу, чтобы образовать четырехмерное нормальное гиперболическое дифференцируемое многообразие. Не все эти довольно далеко идущие предположения с необходимостью следует переносить в квантовую теорию. Например, в некоторых формулировках квантовой ОТО отказались от четырехмерности (см. [92], стр. 495—500).

Принцип эквивалентности подразумевает, что все локальные эффекты гравитационного поля можно исключить, описывая всю физику в окрестности какого-либо события в локальной свободно падающей системе координат. В малом, т. е. в окрестности первого порядка, пространство-время ОТО аналогично пространству-времени Минковского специальной теории относительности, даже при наличии гравитационного поля. (Этот, локально более минковский, чем евклидов, характер пространства-времени называется «нормально гиперболическим».) Отклонения от свойств пространства-времени Минковского появляются во втором порядке, и они отличают истинное гравитационное поле от инерциального, которое можно полностью устранить. Математически эти отклонения являются мерой кривизны пространства-времени; следовательно, истинные гравитационные поля связаны с кривизной пространства-времени, а уравнения гравитационного поля приобретают чисто геометрическую форму соотношений между значениями кривизны.

Для определения понятий кривизны пространства-времени, параллельного переноса и ковариантного дифференцирования требуется ряд рассуждений, которые можно найти в любом стандартном учебнике по дифференциальной геометрии [13, 46, 49], а в краткой форме — в руководствах по ОТО [1, 2, 60, 69]. Для читателей, не знакомых с этими понятиями, мы даем ниже краткий набросок.

Метрика. Поскольку постулируется, что пространство-время является четырехмерным многообразием, каждой точке пространства-времени можно сопоставить четырехмерное касательное векторное пространство с обычными алгебраическими свойствами (элементы которого удобно рассматривать как производные по направлениям). Таким образом, искривленное пространство-время может нести физические векторные и тензорные поля. Однако гравитационно-инерциальное поле определяется заданием (нормально гиперболического) скалярного произведения (метрики) в каждом локальном векторном пространстве. Непосредственный физический смысл имеет не метрика, а траектории свободно падающих пробных частиц, которые, согласно принципу эквивалентности, однозначны в том смысле, что не зависят от состава частиц. В ОТО гравитационное поле геометризуется путем отождествления этих траекторий с геодезическими подходящим образом искривленного пространства-времени. Можно показать, что метрика определяет геодезические, и наоборот. Поэтому метрику можно представлять себе как краткую сводку локальных свойств траекторий пробных частиц. Для математического определения точек, векторов и метрики многообразия удобно выбрать систему координат (или, если необходимо, атлас координатных окрестностей), т. е. совокупность четырех функций x^μ ($\mu = 0, 1, 2, 3$) на многообразии, однозначно описывающих точки, и базис $\mathbf{e}_\mu$ векторного пространства в каждой точке пространства-времени. Обычный (голономный) набор базисных векторов есть набор производных по направлению в четырех координатных направлениях: $\mathbf{e}_\mu = \partial/\partial x^\mu$. Метрика тогда определяется заданием десяти скалярных произведений базисных векторов как функций пространства-времени (т. е. как функций координат со стандартными трансформационными свойствами относительно изменений координат), так как линейность позволяет вычислить скалярное произведение любой пары векторов, разложенных по базису:

$$\mathbf{A} = A^\mu \mathbf{e}_\mu, \quad \mathbf{B} = B^\nu \mathbf{e}_\nu, \quad g_{\mu\nu}(x^\mu) = \mathbf{e}_\mu \cdot \mathbf{e}_\nu,$$
$$\mathbf{A} \cdot \mathbf{B} = A^\mu B^\nu (\mathbf{e}_\mu \cdot \mathbf{e}_\nu) = A^\mu B^\nu g_{\mu\nu}.$$

Здесь мы подразумеваем суммирование по повторяющимся индексам от нуля до трех.

Параллельный перенос. В пространстве-времени с заданной метрикой нет однозначного понятия «параллельности» для векторов в касательных пространствах, принадлежащих двум различным точкам. Однако если дополнительно задана кривая, соединяющая эти точки, то существует понятие параллельного переноса векторов вдоль этой кривой от одной конечной точки к другой, которое однозначно, если мы требуем сохранения скалярных произведений и определенных линейных свойств. Эти свойства, а также сам параллельный перенос можно определить через «инфинитезимальный параллельный перенос» или через ковариантную производную (по направлению), связанную соответствующим образом с параллельным переносом. Пусть s — аффинный параметр вдоль кривой, $\mathbf{A}(s)$ — произвольное векторное поле, заданное на кривой, $\mathbf{A}^{P}(s)$ — параллельный перенос вектора $\mathbf{A}(0)$ на расстояние s вдоль кривой и $\mathbf{T} = d/ds$ — касательная к кривой; определим ковариантную производную $\mathbf{A}$ в направлении $\mathbf{T}$ при $s = 0$ как

$$D_{\mathbf{T}}\mathbf{A} = \lim_{s\to 0} \frac{1}{s}\{\mathbf{A}(s) - \mathbf{A}^{(P)}(s)\}.$$

Обратно, если известна ковариантная производная (более частный случай), то параллельный перенос вектора, а следовательно, и всех тензоров можно определить с помощью интегрирования.

Величина $D_{\mathbf{T}}\mathbf{A}$ должна быть функцией, линейной по $\mathbf{T}$ и $\mathbf{A}$ (последнее в смысле дифференцирования). Поэтому для однозначного определения ковариантной производной необходимо задать только производные всех базисных векторов по всем базисным направлениям. Компоненты этих базисных производных являются коэффициентами вращения Риччи $\gamma^{\varkappa}_{\mu\nu}$:

$$D_{\mathbf{e}_\mu}\mathbf{e}_\nu = \gamma^{\varkappa}_{\mu\nu}\mathbf{e}_\varkappa.$$

В голономном базисе они называются «символами Кристоффеля» и обозначаются $\Gamma^{\varkappa}_{\mu\nu}$. Их можно непосредственно выразить через обыкновенные производные от метрических коэффициентов:

$$\Gamma^{\varkappa}_{\mu\nu} = \frac{1}{2} g^{\varkappa\lambda}(g_{\lambda\mu,\nu} + g_{\lambda\nu,\mu} - g_{\mu\nu,\lambda}). \tag{3.1}$$

С помощью этих коэффициентов компоненты ковариантной производной можно записать полностью через обыкновенные производные по координатам (обозначаемые запятыми, например $A^{\mu}_{,\nu} = \partial A^{\mu}/\partial x^{\nu}$):

$$D_{\mathbf{T}}\mathbf{A} = (A^{\mu}_{;\nu}T^{\nu})\,\mathbf{e}_\mu,$$

где

$$\mathbf{A} = A^{\mu}\mathbf{e}_\mu, \quad \mathbf{T} = T^{\nu}\mathbf{e}_\nu$$

и

$$A^{\mu}_{;\nu} = A^{\mu}_{,\nu} - \Gamma^{\mu}_{\nu\lambda}A^{\lambda}.$$

Точка с запятой (;) используется для обозначения ковариантной производной в пространстве-времени. Ковариантные производные в трехмерной пространственной геометрии будут обозначаться не точкой с запятой, а косой чертой (/).

Геодезические. Геодезические определяются через параллельный перенос (и, следовательно, через метрику) как кривые, касательные к которым остаются параллельными в смысле параллельного смещения вдоль кривой. Это условие в дифференциальной форме означает, что ковариантная производная касательного вектора в касательном направлении равна нулю. Пусть кривая описывается с помощью голономных координат уравнениями $x^\mu = X^\mu(t)$. Тогда компоненты касательного вектора $\mathbf{u}$ равны $u^\mu = dX^\mu/dt$, и мы получаем

$$D_{\mathbf{u}}\mathbf{u} = 0 = \left(\frac{d^2X^\mu}{dt^2} + \Gamma^\mu_{\varkappa\lambda}\frac{dX^\lambda}{dt}\frac{dX^\varkappa}{dt}\right)\mathbf{e}_\mu. \tag{3.2}$$

Это условие также следует из вариационного принципа, согласно которому длина кривой между фиксированными конечными точками принимает экстремальное значение. Длина кривой σ определяется через длину ее касательного вектора $\mathbf{u}$ как

$$\sigma_1^2 = \int\limits_1^2 (\mathbf{u}\cdot\mathbf{u})^{1/2}\,dt.$$

Подынтегральное выражение обычно обозначают ds (хотя это не полный дифференциал, как, казалось бы, указывает обозначение); в голономной системе мы, следовательно, имеем

$$\sigma_1^2 = \int\limits_1^2 ds,$$

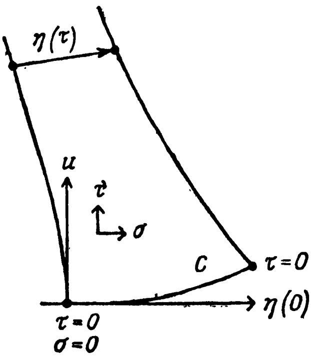
где

$$ds^2 = g_{\mu\nu}\frac{dX^\mu}{dt}\frac{dX^\nu}{dt}dt^2 = g_{\mu\nu}\,dX^\mu\,dX^\nu.$$

Вариационный принцип для геодезических в этом случае имеет простой вид

$$\delta\sigma_1^2 = \delta\int\limits_1^2 ds = 0.$$

Кривизна. Линейные свойства ковариантного дифференцирования интуитивно означают, что параллельный перенос данного вектора в окрестности первого порядка малости порождает локальное поле векторов, обладающих всеми свойствами векторов в обычном пространстве Минковского. Кривизна служит мерой отклонений от этого поведения, обнаруживаемых во втором порядке, таких, как невозможность возвращения вектора к своему начальному направлению после параллельного переноса вдоль замкнутого контура, отклонение суммы углов геодезического многоугольника от ее значения в плоском пространстве или невозможность замкнуть фигуру, построенную из длин и углов многоугольников плоского пространства, некоммутативность ковариантного дифференцирования, нарушение эквидистантности параллельных геодезических. Рассмотрим последнее свойство, так как его можно непосредственно интерпретировать



Фиг. 1. Интервал между времениподобными геодезическими, параллельными вначале на кривой C, измеряется вектором $\eta(\tau) \equiv \partial/\partial\sigma \,|_{\tau=\mathrm{const}}$.

посредством траекторий пробных частиц. Пусть некоторая (времениподобная) геодезическая обладает аффинным параметром τ и единичным касательным вектором u^μ, и пусть C — (пространственноподобная) геодезическая, пересекающая первую при $\tau = 0$, с аффинным параметром σ и единичным касательным вектором η^μ. Построим далее геодезические, начинающиеся на C, начальные направления которых параллельны смещениям u. Точки на этих геодезических естественно обозначить параметрами (σ, τ). Пусть $\eta_\mu(\tau)$ — компоненты вектора $\partial/\partial\sigma\,|_{\tau=\mathrm{const}}$, связывающего точки с равным τ на данной и параллельной ей геодезических. В плоском пространстве вектор $\eta(\tau)$ постоянен, так как параллельные геодезические остаются эквидистантными. В кривом пространстве $(d\eta/d\tau)(0) = 0$, т. е. геодезические, параллельные вначале, остаются эквидистантными в окрестности первого порядка, а ускорение

$$\left.\frac{d^2\eta^\mu}{d\tau^2}\right|_{\tau=0} = - R^\mu_{\nu\varkappa\lambda} u^\nu \eta^\varkappa u^\lambda \tag{3.3}$$

является мерой кривизны (фиг. 1). Здесь $R^\mu_{\nu\varkappa\lambda}$ — компоненты тензора кривизны Римана — Кристоффеля. Таким образом, этот тензор можно рассматривать как характеристику одного свойства траекторий пробных частиц, а именно их относительного ускорения. Неудивительно поэтому, что эти компоненты можно выразить через компоненты метрики; фактически их можно полностью записать с помощью символов Кристоффеля:

$$R^{\mu}_{\nu\varkappa\lambda} = \Gamma^{\mu}_{\nu\lambda\varkappa} - \Gamma^{\mu}_{\lambda\varkappa,\,\lambda} + \Gamma^{\mu}_{\alpha\varkappa}\Gamma^{\alpha}_{\nu\lambda} - \Gamma^{\mu}_{\alpha\lambda}\Gamma^{\alpha}_{\nu\varkappa}. \tag{3.4а}$$

Вспоминая уравнение (3.1), видим, что $R^{\mu}_{\nu\varkappa\lambda}$ имеет второй дифференциальный порядок по компонентам метрики.

При помощи операции взятия следа («свертки») можно образовать тензоры низших рангов из тензора Римана — Кристоффеля четвертого ранга, компоненты которых будут свернутыми суммами тензора $R^{\mu}_{\nu\varkappa\lambda}$:

$$R_{\mu\nu} = R^{\alpha}_{\mu\alpha\nu} \quad \text{(тензор Риччи)}, \tag{3.4б}$$

$$R = g^{\alpha\beta}R_{\alpha\beta} \quad \text{(скалярная кривизна)}. \tag{3.4в}$$

Эти суммы в теории гравитации соответствуют действующим на $g_{\mu\nu}$ дифференциальным операторам типа лапласиана, используемого в других теориях поля.

Для квантования, а также для более четкого понимания классической теории часто целесообразно разделять пространство и время и соответственно расщеплять четырехмерные тензоры. Подобное описание мы будем называть $(3+1)$-мерным (следуя Уилеру). Предположим, что задана трехмерная пространственноподобная гиперповерхность постоянного t. Пусть базисные векторы выбраны так, что три из них касательны к гиперповерхности, т. е. выбран голономный базис $e_i = \partial/\partial x^i\,|_{t=\text{const}}$. (Здесь и в дальнейшем латинские индексы принимают значения от 1 до 3 и соответствуют трехмерным или $(3+1)$-мерным объектам, а греческие изменяются от 0 до 3 и обозначают пространство-время или четырехмерные объекты. В случае необходимости используется также индекс 4 слева сверху для пространственно-временных величин, 4-векторы обозначаются полужирным шрифтом, а трехмерные векторы — стрелками сверху.) Тогда пространственно-временная метрика индуцирует естественную метрику на гиперповерхности:

$$g_{ij} = \vec{e}_i \cdot \vec{e}_j = \mathbf{e}_i \cdot \mathbf{e}_j = {}^4g_{ij}. \tag{3.5}$$

Следовательно, в этом базисе пространственные компоненты пространственно-временной метрики являются также компонентами метрики на пространственной гиперповерхности. Остальные компоненты $g_{\mu\nu}$ имеют простой смысл, если мы используем в качестве времениподобного базисного вектора голономный вектор $\mathbf{e}_0 = \partial/\partial t$. Его можно рассматривать как вектор, соединяющий точки с одинаковыми пространственными координатами на одной какой-либо гиперповерхности и на следующей за ней гиперповерхности. Он не обязательно перпендикулярен гиперповерхности; действительно, три «функции сдвига»

$$N_i = \mathbf{e}_0 \cdot \mathbf{e}_i = g_{0i}$$

определяют величины его проекций на гиперповерхность, а его длина задается оставшейся компонентой метрики $\mathbf{e}_0 \cdot \mathbf{e}_0 = g_{00}$. Часто более удобно использовать длину N его проекции на единичную нормаль к гиперповерхности (функцию смещения) и его проекцию на гиперповерхность — 3-вектор $\vec{N}$. Определим g^{ij} как матрицу, обратную метрике g_{ij}. Имеются следующие соотношения между функциями сдвига и смещения, пространственно-временной компонентой метрики g_{00} и компонентами N^i вектора сдвига $\vec{N}$:

$$\vec{N} = N^i \vec{e}_i, \quad N^i = g^{ij} N_j, \quad g_{00} = \vec{N}^2 - N^2.$$

Чтобы завершить список геометрических $(3+1)$-объектов, которые нам понадобятся, определим тензор внешней кривизны

$$K_{ij} = -n_{i;\,j}, \tag{3.6}$$

где n^μ — компоненты единичного 4-вектора, нормального к гиперповерхности, а точка с запятой (;) обозначает пространственно-временную ковариантную производную. Внешняя кривизна описывает погружение гиперповерхности в окружающее пространство, измеряя, как изменяется нормаль при ее параллельном переносе по гиперповерхности. Нетрудно показать [92], что тензор K_{ij} связан с g_{ij}, N и N_i соотношением

$$K_{ij} = \frac{1}{2N} (N_{i|j} + N_{j|i} - \dot{g}_{ij}), \tag{3.7}$$

где точка обозначает $\partial/\partial t$. В теории Эйнштейна более непосредственно, чем тензор K_{ij}, появляется тензор плотности

$$\pi^{ij} = g^{1/2} (g^{ij} K - K^{ij}), \tag{3.8}$$

где g — детерминант g_{ij}, а $K = \operatorname{Sp} K = g^{ij} K_{ij}$.

2. Принцип действия, уравнения поля и тождества Бианки

Для свободного гравитационного поля действие строится из скалярной кривизны:

$$I = \int {}^4R\,({}^4g)^{1/2}\,d^4x, \qquad (3.9)$$

где $({}^4g)^{1/2}\,d^4x$—инвариантный элемент объема. Из метрики можно построить и другие скалярные плотности действия, но именно эта форма принципа действия является единственной, дающей уравнения поля, линейные по вторым производным от $g_{\mu\nu}$, и сводящейся к ньютоновскому случаю в пределе слабых полей. Если присутствуют другие поля, отличные от гравитационного, то их действие, обобщенное на случай кривого пространства, необходимо просто добавить к приведенному выше; взаимодействие полей полностью учитывается при обобщении действия на случай искривленного пространства, и введения каких-либо специальных членов взаимодействия не требуется (принцип минимальности взаимодействия).

Для получения уравнений поля действие (3.9) нужно проварьировать по $g_{\mu\nu}$. Оказывается также, что эквивалентные результаты получаются, если не учитывать уравнение (3.1) и варьировать символы Кристоффеля и метрические коэффициенты в уравнении (3.9) независимо (вариация Палатини). Последняя вариация легко вычисляется, поскольку $g_{\mu\nu}$ при этом не дифференцируется; в результате получается

$$G^{\mu\nu} \equiv R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R = 0. \qquad (3.10)$$

[В случае присутствия материальных полей правая часть уравнения (3.10) не равна нулю и проявляется источник в виде тензора энергии-импульса материи $T^{\mu\nu} = \delta\mathscr{L}_{\text{мат}}/\delta g_{\mu\nu}$ [1]).] Вариация символов Кристоффеля в (3.4) должна давать, с другой стороны, ковариантные дифференциальные уравнения первого порядка по $g_{\mu\nu}$ [так как в интеграле действия (3.9) содержатся Γ и их первые производные], а именно

$$g_{\mu\nu;\,\varkappa} = 0. \qquad (3.11)$$

---

[1]) Утверждение об эквивалентности варьирования по $g_{\mu\nu}$ (g-варьирования) и варьирования Палатини (P-варьирования) справедливо только при игнорировании лагранжевой плотности внешнего материального поля. Если учесть наличие последнего, то P-варьирование приводит к существенно новой теории, связанной с появлением тензора кручения [95—98]. — *Прим. перев.*

Эти уравнения являются точными условиями того, что Г представляет собой выделенную связь, определяемую метрикой, и они эквивалентны уравнениям (3.1). Если только $g_{\mu\nu}$ варьируются независимо, а вариация Г определяется выражением (3.1), то опять получаются уравнения поля (3.10), так как дополнительные члены из зависимой вариации Г являются линейными комбинациями обращающихся в нуль выражений (3.11).

На первый взгляд кажется, что уравнения поля состоят из десяти уравнений для десяти независимых компонент метрического тензора; однако четыре «свернутых тождества Бианки»

$$G^{\nu}_{\mu;\,\nu} = 0 \tag{3.12}$$

[правильность которых можно проверить непосредственно с помощью уравнений (3.1) и (3.4)] означают, что не все десять уравнений независимы. Действительно, шесть уравнений Эйнштейна

$$G^{l}_{i} = 0$$

вместе с тождествами Бианки дают

$$G^{0}_{\mu,\,0} = 0,$$

так что четыре уравнения Эйнштейна

$$G^{0}_{\mu} = 0 \tag{3.13}$$

удовлетворяются везде в пространстве-времени, если они удовлетворяются на какой-нибудь одной гиперповерхности. Эти четыре уравнения не добавляют ничего в определение развития во времени тензора $g_{\mu\nu}$, ограничивая только выбор начальных значений для $g_{\mu\nu}$ и $\dot{g}_{\mu\nu}$ [1]). Следовательно, временнóе развитие $g_{\mu\nu}$ определяется только с точностью до четырех произвольных пространственно-временных функций, которые можно отождествить с пространственно-временными координатами. Четыре уравнения (3.13) называются уравнениями начальных значений или просто связями. В настоящее время выяснено, что подобные связи всегда возникают в общековариантных теориях [10, 22].

---

[1]) Чтобы убедиться, что G^{0}_{μ} зависит только от $g_{\mu\nu}$ и $\dot{g}_{\mu\nu}$ заметим, что некоторые члены в сумме (3.12) не содержат производных по времени выше второго порядка. Так как (3.12) — тождество, то отсюда следует, что никакой член не может содержать производные по времени выше второго порядка, поэтому G^{0}_{μ} не может содержать производных по времени выше первого порядка.

3. Преимущества разделения времени и пространства

До недавнего времени необходимость пространственно-временной ковариантности в ОТО не подвергалась сомнению. Но, как оказалось, явное разделение пространства и времени в $(3+1)$-ковариантном формализме имеет большие формальные и принципиальные преимущества в ОТО; установление этого факта имеет очень важное значение. Здесь мы продемонстрируем некоторые из этих преимуществ, сначала чисто формальные, а затем те, которые имеют физическую интерпретацию, и, наконец, приведем удивительно простую и ясную форму принципа действия в $(3+1)$-формализме.

$(3+1)$-Подход оперирует с «внутренними» тензорами, образованными из положительно определенной 3-метрики на гиперповерхности, задаваемой уравнением (3.5), и «внешними» тензорами, образованными из функций сдвига и смещения, а также внешней кривизной, определяемой уравнениями (3.6) и (3.7). Все геометрические пространственно-временные величины должны быть выражены через эти более простые объекты. Например, компоненты пространственно-временной кривизны определяются выражениями

$$R^{0a}_{0b} = -N^{-1}\dot{K}^a_b + K^a_c K^c_b + N^{-1}N^r\left(K^a_{b,r} + K^a_{r/b} - K^a_{b/r}\right), \qquad (3.14)$$

$$R^{0a}_{pq} = N^{-1}\left(K^a_{p/q} - K^a_{q/p}\right), \qquad (3.15)$$

$$\begin{aligned} {}^4R^{ab}_{cd} = R^{ab}_{cd} + K^a_c K^b_d - K^a_d K^b_c + \\ + N^{-1}\left\{N^b\left(K^a_{c/d} - K^a_{d/c}\right) - N^a\left(K^b_{c/d} - K^b_{d/c}\right)\right\}. \end{aligned} \qquad (3.16)$$

Двадцать независимых компонент $R^{\mu}_{\nu\alpha\beta}$ заменяются двенадцатью компонентами тензоров внутренней и внешней кривизны и четырьмя функциями смещения и сдвига. Уменьшение числа компонент происходит в результате того, что компоненты R^{0a}_{pq} связаны с пространственными производными от K^a_b, поэтому компоненты внутренней кривизны служат потенциалами для некоторых компонент пространственно-временной кривизны.

Полезность для физической интерпретации разделения в теории относительности пространства и времени хорошо известна [69]. В реальных экспериментах редко измеряют одновременно и пространственноподобные и времениподобные компоненты какой-либо величины. Обычно разделяют энергию и импульс частицы, электрические и магнитные компоненты максвелловского поля, пространственные и временные компоненты пространственно-временного смещения, так как для определения каждой такой пары необходимы два отдельных эксперимента. Матрицы внутренней и внешней кривизны также измеряются посредством экспериментов разного типа.

Рассмотрим идеализированную ситуацию гравитационного измерения, в которой наблюдатель измеряет относительное ускорение двух свободных пробных тел, на мгновение покоящихся относительно друг друга. (В реальных экспериментах пробные тела не свободны, но обычно достаточно учесть поправку на силы связи при помощи простой ньютоновской механики.) Пусть четыре координаты первого тела в момент τ_1 его собственного времени равны $x(\tau_1)$, и пусть $y(\tau_2)$ — координаты другого тела в момент τ_2 его собственного времени. При равных собственных временах разность координат тел составляет

$$\eta^\mu(\tau) = x^\mu(\tau) - y^\mu(\tau).$$

Общая 4-скорость в момент измерения равна

$$u^\mu = \frac{dx^\mu}{d\tau} = \frac{dy^\mu}{d\tau}.$$

Относительное ускорение, определяемое выражением

$$a^\mu \equiv \left(\frac{D^2}{D\tau^2}\right)\eta^\mu \equiv (\eta^\mu_{;\nu} u^\nu)_{;\varkappa} u^\varkappa,$$

хорошо аппроксимируется с помощью уравнения геодезического отклонения (3.3), если расстояние между частицами мало по сравнению с радиусом пространственно-временной кривизны

$$a^\mu = - R^\mu_{\beta\alpha\delta} u^\beta \eta^\alpha u^\delta. \tag{3.17}$$

Представим это уравнение в $(3+1)$-форме применительно к простому случаю, когда пробные тела на мгновение находятся в покое в локальной инерциальной системе отсчета гиперповерхности $x^0 = 0$ (т. е. в локальной, ортогональной времени системе отсчета, в которой события на гиперповерхности одновременны, т. е. $N = 1$, $N_i = 0$). В этой системе отсчета уравнения (3.14) и (3.17) дают

$$a^0 = 0$$

(часы остаются синхронизированными до второго порядка по $\Delta\tau$),

$$\vec{a} = ([\dot{K}] + [K]^2)\,\vec{\eta}. \tag{3.18}$$

Матрица внешней кривизны $[K] = [K^a_b]$, очевидно, определяет тензор силы, действующий на небольшую группу тел, покоящихся в системе отсчета гиперповерхности. Она аналогична электрической компоненте максвелловского поля, которая описывает силы, действующие на покоящиеся заряды. Отсутствие компонент внутренней кривизны R^{ab}_{cd} в уравнении (3.18) и появление этих компонент в более общем уравнении (3.17) всего лишь в виде коэффициентов при пространственноподобной скорости $\vec{u}$ указывают на то, что измерение внутренней кривизны возможно только тогда, когда пробные тела движутся относительно системы отсчета гиперповерхности. Тензор внутренней кривизны аналогичен магнитной компоненте максвелловского поля, описывающей силы, действующие на токи.

Поскольку $[K]$ является симметричной 3×3-матрицей, которую можно выразить через определяемые ею квадратичные поверхности, можно пойти дальше в разбиении кривизны на физически отдельные элементы. Можно разложить матрицу на изотропную и бесследовую части и сопоставить первую с космологическими эффектами, а вторую — с такими локальными гравитационными явлениями, как ньютоновская гравитация и гравитационные волны. Как отметил Уилер [92], бесследовая часть может быть далее проанализирована с помощью параметра интенсивности и «фактора формы», являющегося мерой отклонения от осевой симметрии. Аналогичному анализу может быть подвергнута «матрица внутренней кривизны»

$$G^a_b \equiv R^a_b - \frac{1}{2} R\delta^a_b.$$

Все сказанное выше о важности $(3+1)$-формализма для ОТО могло быть высказано гораздо раньше. Однако это не было бы убедительным, поскольку считалось, что пространственно-временное описание является более подходящим, обеспечивая более простые геометрические формы для принципа действия и уравнений поля. Современный интерес к $(3+1)$-формализму возник тогда, когда нашли, что этот формализм действительно упрощает принцип действия и проблему начальных значений в ОТО. Байерлейн и др. [7] установили, что геометрия гиперповерхности (внутренняя или внешняя) является величиной, которая может быть свободно задана в начальный и конечный моменты времени во вселенной с замкнутыми пространственными сечениями. Арновит, Дезер и Мизнер (АДМ) [5] показали, что действие в ОТО можно представить в изящной форме

$$^{8}I' = \int dt \int d^3x \, (\dot{g}_{ij}\pi^{ij} - N_\mu R^\mu), \tag{3.19}$$

где

$$-R^i = -2g^{1/2}NG^{0i} = 2\pi^{ij}_{|j}$$

и

$$-R^0 = 2g^{1/2}N^2G^{00} = g^{1/2}(R + K^2 - \mathrm{Sp}\,[K]^2).$$

В табл. 1 данная форма действия сравнивается с некоторыми из наиболее распространенных альтернативных форм (они отличаются интегралами по поверхности и полными производными по времени). В настоящее время считается, что $(3+1)$-формализм во многих отношениях лучше подходит для анализа динамики свободного гравитационного поля, чем пространственно-временной подход.

ТАБЛИЦА 1

Функционалы действия ОТО

Определения:

$$^4I \equiv \int {}^4\sigma\, {}^4R,$$

$$I_{\Gamma\Gamma} \equiv \int d^4x\,(-{}^4g)^{1/2}\,(\Gamma^{\mu}_{\nu\lambda}\Gamma^{\lambda}_{\varkappa\mu} - \Gamma^{\lambda}_{\nu\varkappa}\Gamma^{\mu}_{\lambda\mu})g^{\nu\varkappa},$$

$$^3I \equiv \int dt \int d^3x\,(-g_{ij}\dot{\pi}^{ij} - N_{\mu}R^{\mu}),$$

$$^3I' \equiv \int dt \int d^3x\,(\dot{g}_{ij}\pi^{ij} - N_{\mu}R^{\mu}).$$

Разбиение, использующее 3-геометрию:

$$^4I = \int dt \int \sigma\{-2\dot{K} + N(R + K^2 + \mathrm{Sp}\,\mathbf{K}^2) + 2N^bK_{/b}\}.$$

Соотношения, связывающие различные выражения для действия:

$$I_{\Gamma\Gamma} = {}^3I' - \int dt \int \sigma\,\{2N^iK + N(g^{sb}\Gamma^i_{sb} - g^{si}\Gamma^b_{sb})\}_{/i},$$

$$^3I = {}^4I + 2\int dt \int \sigma\,(N_jK^{ij})_{/i},$$

$$^3I' = {}^3I + 2\int dt \int \sigma\left\{g^{-1/2}\frac{\partial}{\partial t}(g^{1/2}K)\right\}.$$

§ 4. КАНОНИЧЕСКАЯ ФОРМА ОБЩЕЙ ТЕОРИИ ОТНОСИТЕЛЬНОСТИ

1. Природа канонического формализма

Стандартная процедура приведения теории с динамическими переменными q^A и лагранжианом $L(q, \dot{q})$ к канонической форме требует определения импульсов

$$p_A = \frac{\partial L}{\partial \dot{q}^A}, \tag{4.1}$$

выражения через них скоростей $\dot{q}^A$, а затем определения гамильтонова функционала

$$H(q, p) = \sum_A \dot{q}^A p_A - L(q, p). \tag{4.2}$$

Эта процедура применима лишь в том случае, если можно разрешить (4.1) относительно скоростей $\dot{q}^A$ [т. е. если лагранжиан несингулярен: $\det(\partial^2 L/\partial\dot{q}^A\partial\dot{q}^B) \neq 0$]. Однако обычный, содержащий лишь первые производные лагранжиан ОТО $I_{\text{ГГ}}$ (табл. 1) сингулярен — некоторые из определяющих импульсы уравнений совсем не содержат производных по времени.

Эта трудность возникает вследствие того, что эйнштейновская формулировка ОТО содержит большее число переменных, нежели число физических степеней свободы: эйнштейновские полевые переменные (компоненты метрического тензора пространства-времени $g_{\mu\nu}$) содержат информацию не только о динамических переменных (описывающих полевые степени свободы), но также и о координатных системах. Представим теперь, как сформулировал бы теорию какой-нибудь «всезнающий» теоретик: каждую пространственно-временную точку P он отметил бы четырьмя определенными, имеющими физический смысл координатами, а поле охарактеризовал бы двумя динамическими переменными $q^A(P)$ и сопряженными им импульсами $p_A(P)$. В настоящее время только такой «всезнающий» теоретик знает, как разделить эйнштейновские переменные на истинные динамические переменные и на «реальные» координаты. В лагранжиане появляются только динамические переменные, и лишь они и сопряженные им импульсы являются независимыми начальными величинами; следовательно, сопряженные координатам «импульсы», которые можно получить, если слепо следовать правилу (4.1), имеют нулевые значения. В эйнштейновском варианте теории появляются четыре уравнения связи

$$R^\mu = 0 \quad (\mu = 0, 1, 2, 3),$$

связывающие переменные и импульсы в каждой пространственно-временной точке. Ниже мы увидим, что четыре из эйнштейновских полевых переменных являются в сущности множителями Лагранжа, служащими для включения этих координатно-сопряженных [1]) уравнений связи в вариационный принцип.

[1]) Легко показать, что R^μ, определенные в (3.19), являются действительно импульсами, сопряженными пространственно-временным координатам. Используем тот факт, что $\delta g_{\mu 0} = \delta\dot{x}_\mu + \delta^0_\mu \delta\dot{x}_0$ при варьировании по $\dot{x}^\mu$ без изменения x^μ на данной гиперповерхности постоянных x. Здесь $\delta x_\mu = g_{\mu\nu}\delta x^\nu$ и $\delta\dot{x}_\mu \equiv (\delta x_\mu)_{;0}$.

Сингулярность эйнштейновского лагранжиана следует из трактовки этих множителей как переменных поля.

2. Функции смещения и сдвига как лагранжевы множители для учета связей

Варьирование $(3+1)$-формы действия (3.19) по $\dot{g}_{ij}$ определяет π^{ij} как импульсы, сопряженные g_{ij}. Вариация по g_{ij} дает шесть динамических уравнений Эйнштейна, которые можно записать в гамильтоновой форме, если выразить $\dot{\pi}^{ij}$ как явные функции от π^{ij} и g_{ij} и их пространственных производных (см. [5]). Вариация по N_μ дает уравнения

$$R^\mu = 0 \tag{4.3}$$

[см. уравнение (3.19)], являющиеся уравнениями связи, так как они не содержат производных по времени от g_{ij} и π^{ij}. Поскольку функции смещения и сдвига N_μ никогда, даже при определении π^{ij}, не дифференцируются по времени, имеет смысл отождествить их с множителями Лагранжа, служащими для учета в вариационном принципе координатно-сопряженных уравнений связи.

Возвращаясь к стандартному подходу к каноническому формализму, теперь можно выяснить, где он неприменим. Импульсы, сопряженные функциям смещения и сдвига, определяются как

$$\pi^\mu = \frac{\delta L}{\delta \dot{N}_\mu}$$

и обращаются в нуль, так как $\dot{N}_\mu$ никогда не входят в лагранжиан:

$$\pi^\mu = 0. \tag{4.4}$$

Эти уравнения, называемые первичными уравнениями связи, до сих пор не удавалось записать в столь простой форме вследствие неудачи при приведении действия к $(3+1)$-форме (3.19) [24].

Развивая дальше стандартный канонический метод, строим гамильтониан

$$H = \int d^3x \, (\dot{N}_\mu \pi^\mu + N_\mu R^\mu). \tag{4.5}$$

Затем переопределяем π^{ij} и динамические уравнения Эйнштейна в канонической форме:

$$\dot{g}_{ij} = \frac{\delta H}{\delta \pi^{ij}}, \tag{4.6}$$

$$\dot{\pi}^{ij} = -\frac{\delta H}{\delta g_{ij}}. \tag{4.7}$$

Получаем также уравнения

$$\dot{N}_\mu = \frac{\delta H}{\delta \pi^\mu} \equiv \dot{N}_\mu, \tag{4.8}$$

$$\dot{\pi}^\mu = -\frac{\delta H}{\delta N_\mu} \equiv R^\mu. \tag{4.9}$$

Уравнение (4.8) показывает, что $N_\mu(t)$ совершенно произвольны, что отражает произвольность четырех координат, используемых для описания пространства-времени. Мы также замечаем, что уравнение (4.9) само по себе не означает обращения в нуль координатно-сопряженных уравнений связи для R^μ. Поэтому условия связи $R^\mu = 0$ необходимо наложить отдельно, чтобы обеспечить совместность с первичными уравнениями связи (4.4). Вследствие этого обстоятельства исторически сложилось так, что уравнения (4.3) назвали «вторичными уравнениями связи». Дирак [29, 30] показал, что появление вторичных связей, которые необходимо наложить на решения уравнений Гамильтона для обеспечения совместности, является общим для всех общековариантных теорий поля. В этом смысле ОТО является действительно удачной теорией, так как тождества Бианки обеспечивают сохранение связей R^μ. В общем случае может возникнуть необходимость наложить дальнейшие вторичные связи для получения непротиворечивой теории.

Необычной особенностью канонического гамильтониана (4.5) является тот факт, что он равен нулю для любого поля, удовлетворяющего первичным и вторичным связям. Если мы отождествим H с полной энергией и вспомним, что для собственных стационарных состояний квантованной системы эволюция во времени описывается множителем $\exp(iEt)$ (в картине Шредингера), то нам придется согласиться с тем, что квантовые волновые функции не могут существенно зависеть от времени. С этой трудностью столкнулись довольно рано и назвали ее проблемой «замороженного формализма» [11, 59, 66]. Этот результат в целом не удивителен, так как в теориях того типа, к которому применима подобная аргументация, время является произвольным параметром. Наиболее глубоко разработанной разновидностью этой точки зрения на квантование является в настоящее время схема квантования Уилера — де Витта, о которой подробнее будет сказано в § 5 и 7.

В открытых моделях вселенной, где время может быть хорошо определено, по крайней мере на пространственной бесконечности, было найдено, что полная энергия гравитационного поля не равна гамильтониану (4.5), отличаясь от него поверхностным интегралом (см. п. 3 данного параграфа). Таким образом, проблема замороженного формализма присуща исключительно замкнутым моделям вселенной.

Канонический формализм, который мы рассматривали (назовем его общековариантным каноническим формализмом), не является единственным способом формулировки ОТО в каноническом виде. Вследствие налагаемых связей (как первичных, так и вторичных) общековариантный формализм не является строго аналогичным общепринятой гамильтоновой теории, в которой все следует из уравнений Гамильтона. Формализм, который мы рассматриваем, не использует того факта, что N_μ являются в действительности множителями Лагранжа. Величины N_μ рассматриваются наравне с g_{ij}; их особая природа проявляется лишь в гамильтоновых уравнениях (4.8) и (4.9), которые выводятся с их помощью. Теперь мы рассмотрим альтернативный подход к канонической форме ОТО, в котором в явном виде используются N_μ в качестве множителей Лагранжа и отвергаются уравнения движения в обычной канонической форме.

Для получения обычной канонической формы ОТО необходимо в сущности решить уравнения связей до обсуждения принципа действия. Затем действие

$$\int dtL_c = \int dt \int \sigma\, (\dot{g}_{ij}\pi^{ij}) \tag{4.10}$$

варьируется по тем величинам, которые могут быть произвольными на историях, удовлетворяющих уравнениям связи. Байерлейн и др. [7] выяснили, что этими величинами, по-видимому, являются g_{ij}. Они нашли, что для произвольной истории g_{ij} (g_{ij} как функции пространственных и временных координат) всегда можно найти функции смещения и сдвига N_μ, которые разрешают уравнения связей R^μ. Если g_{ij} описывают замкнутую вселенную, то разумно согласиться, что величины N_μ, которые удовлетворяют эллиптическим дифференциальным уравнениям, единственны с точностью до тривиальных координатных преобразований [1]). Для открытых вселенных единственность не была доказана. Возможно, что в случае открытой геометрии знание величин g_{ij} должно быть дополнено заданием граничных условий для функций смещения и сдвига.

---

[1]) Величина N зависит от определения t в том случае, когда $\dot{g}_{ij}$ оказываются равными нулю. Сдвиг $\vec{N}$ может быть определен только с точностью до вектора Киллинга, если последний существует.

В обычном каноническом формализме импульсы π^{ij} не являются сопряженными g_{ij}, так как π^{ij} в (4.10) явно зависят от g_{ij} через решенные уравнения связей. Если построить, согласно обычному определению, прежние импульсы p^{ij}, сопряженные g_{ij}, то обнаружится, что не все p^{ij} являются независимыми. Легче всего понять это, если учесть, что любые вариации

$$\delta\dot{g}_{ij} = v_{i/j} + v_{j/i} + \phi\dot{g}_{ij}$$

соответствуют координатным изменениям, оставляющим инвариантной реальную геометрическую историю. Импульсы, соответствующие таким вариациям, будут равны нулю, так как действие является инвариантным.

Для перехода к канонической теории необходимо построить канонические переменные, обладающие взаимно однозначным соответствием с геометрическими историями. Одним из подходов является наложение особых условий на g_{ij} и $\dot{g}_{ij}$ с целью фиксировать единым образом представление каждой истории. В качестве канонических переменных берутся те величины, которые могут быть произвольно определены из решений для особых или дополнительных условий. Поскольку имеется шесть g_{ij}, а координатных условий требуется только четыре, то две пространственные функции ϕ_A $(A = 1, 2)$ остаются в качестве канонических переменных. Затем L_c выражают через одни ϕ_A и $\dot{\phi}_A$ и переходят к получению сопряженных импульсов π^A и гамильтониана H_c, как в обычной теории поля. Этот подход разработали АДМ [5], используя технику теории возмущений, для решения уравнений связи и дополнительных условий. В качестве ϕ_A они взяли поперечные бесследовые части поля g_{ij}. Мы вернемся к формализму АДМ в следующем разделе.

Мы видели, что $(3+1)$-форма принципа действия ОТО дает вклад в два различных подхода к канонической форме ОТО — общековариантный и обычный. Основные особенности обоих этих подходов следуют из того факта, что функции смещения и сдвига являются на самом деле множителями Лагранжа, служащими для учета уравнений связи, которые в свою очередь получаются из произвольности координат. Однако в общековариантном подходе этот факт явно не используется и N_μ рассматриваются наравне с g_{ij}. Действительно, общековариантный подход в своей ранней неполной форме предваряет открытие истинной природы N_μ [10, 11, 29]. С другой стороны, обычный канонический формализм есть не что иное, как всестороннее приложение нашей трактовки функций смещения и сдвига и уравнений связи.

3. Асимптотические изометрии и энергия-импульс

В открытых вселенных необходимо дополнить уравнения Эйнштейна, наложив на $g_{\mu\nu}$ граничные условия на бесконечности. Преобразование, которое оставляет граничные условия неизменными, называется асимптотической изометрией. Всестороннее рассмотрение асимптотических граничных условий и изометрий должно включать в себя подробное обсуждение гравитационного излучения (см., например, [80]). Мы ограничимся лишь кратким рассмотрением этих вопросов и вместо этого сконцентрируем свое внимание на центральной роли асимптотических изометрий в обычном каноническом формализме АДМ.

Граничное условие, с которым мы будем иметь дело, налагает на величины на бесконечности требование, чтобы они были «асимптотически плоскими». Первоначально определить это условие можно требованием, чтобы пространство-время было плоским (минковским) везде вне ограниченной области. Асимптотические изометрии, связанные с этим граничным условием, являются неоднородными преобразованиями Лоренца. Но подчиненные такому условию уравнения Эйнштейна имеют только одно решение — всюду плоское пространство Минковского специальной теории относительности. Необходимо ослабить граничное условие, чтобы допустить существование входящего и выходящего гравитационного излучения и ненулевую общую массу вселенной. Пенроуз [71] сформулировал такие граничные условия, выраженные через внутреннюю геометрию гиперповерхности на бесконечности. Как и следовало ожидать, ослабленные граничные условия содержат расширенную группу изометрий — обобщенную группу Бонди — Метцнера — Сакса, которая в качестве подгруппы содержит неоднородную (ортохронную) группу Лоренца [79]. Так как полная группа Бонди — Метцнера еще не до конца изучена, мы будем иметь дело только с лоренцевыми изометриями.

Первым шагом в построении обычного канонического формализма для ОТО является выбор дополнительных условий, которые определяют предпочтительные пространственно-временные координаты T и X^i. В асимптотически плоской геометрии удобно потребовать, чтобы T и X^i стремились на бесконечности к обычным координатам Минковского — Декарта. Если решены уравнения связи R^μ, то смещение N на бесконечности должно стремиться к единице, а сдвиг $\vec{N}$ — к нулю. Рассматривая пространственную метрику как малую флуктуацию плоской декартовой метрики δ_{ij}, АДМ удовлетворяют этим требованиям, по крайней мере во втором порядке своей теории возмущений. Они используют естественное разложение произвольного тензора f_{ij} на поперечную, бесследовую и продольную части:

$$f_{ij} = f_{ij}^{TT} + \frac{1}{2}\delta_{ij} f^{T} - \frac{1}{2}\nabla^{-2} f^{T}_{,ij} + f_{i,j} + f_{j,i},$$

где

$$f_{ii}^{TT} = 0, \quad f_{ij,j}^{TT} = 0,$$

а

$$f^{T} = f_{ii} - \nabla^{-2} f_{ij,ij},$$

и находят в качестве удовлетворительного выбора координатных условий условие «минимального» объема

$$T = -\frac{1}{2}\nabla^{-2}\pi^{T}$$

на гиперповерхности $T = \mathrm{const}$ и гармонические пространственные координатные условия

$$X^{i} = g_{i}.$$

АДМ записывают свои координатные условия в обращенной, или разрешенной форме. В более привычной форме эти координатные условия имеют вид

$$\pi^{T} = 0, \quad \text{или} \quad \mathrm{Sp}\,[K] = 0,$$

и

$$g_{ij,j} = 0.$$

Заметим, что пространственные координаты АДМ являются гармоническими только по отношению к оператору плоского фона ∇^2. Таким образом, они не тождественны известным ковариантным гармоническим координатам, определяемым соотношениями

$$(g^{1/2} g^{ij})_{,j} = 0.$$

Координатные условия АДМ могут быть разрешены, так как они линейны, а геометрия фона является плоской, так что все дифференциальные операции коммутируют. В волновой зоне, или в асимптотической области асимптотически плоского пространства, эта геометрия фона является естественной. Очевидно, что координаты АДМ удовлетворительны в волновой зоне, так как волновая зона определяется как область справедливости низшего порядка теории возмущений и по определению там не может быть серьезных сингулярностей, как реальных, так и координатных.

Используя свой вариант обычной канонической теории, АДМ находят выражения для таких сохраняющихся пространственно-временных величин, как энергия-импульс и полный угловой момент, которые являются поверхностными интегралами, взятыми по всей волновой зоне. Например, полная энергия АДМ дается интегралом по 2-поверхности:

$$E_\infty = \int\limits_\infty N g^{1/2} g^{ij} (g_{ik,\,j} - g_{ij,\,k})\, dS^k \tag{4.11}$$

(который берется по гиперповерхности $t = \text{const}$). Это общая масса, которую удаленный наблюдатель приписал бы вселенной, измеряя ее гравитационное поле и сравнивая его с полем массивной сферы. Поскольку эти выражения справедливы для зоны чистого излучения, результаты теории возмущений строго справедливы и не зависят от выбора координат во внутренних областях.

Чтобы получить общую энергию АДМ из общековариантного гамильтониана (4.5), необходимо дополнить $(3+1)$-лагранжиан (3.19) поверхностным интегралом, эквивалентным в асимптотической области $-E_\infty$. Как показывает табл. 1, это в точности тот поверхностный интеграл, который должен быть добавлен к $(3+1)$-лагранжиану, чтобы получить содержащий первые производные нековариантный лагранжиан $L_{\text{гг}}$.

АДМ утверждают, что их координатные условия могут быть использованы для построения строгой обычной канонической теории гравитации в асимптотически плоском пространстве-времени. В противовес этому утверждению необходимо рассмотреть возможность того, что любое простое множество координатных условий может удовлетворять лишь ограниченному семейству геометрий, в особенности если допускаются большие значения кривизны [80]. Ясно, что пространство-время, для которого координаты АДМ несингулярны, может быть непрерывно отображено на пространство-время Минковского. Если бы могло быть доказано обратное, то формализм АДМ был бы полностью оправдан, так как в настоящее время нет реальной теории, описывающей изменение топологии в ОТО.

§ 5. МЕТОДЫ КВАНТОВАНИЯ

1. Каноническое квантование

Наиболее известный метод квантования представляет состояние[1]) системы вектором $|\psi\rangle$ в гильбертовом простран-

---

[1]) Мы всегда будем рассматривать только чистые состояния. Нам не известны какие-либо удачные попытки более общего рассмотрения состояний с «матрицей плотности» в квантовой ОТО.

стве $\mathscr{H}$. Когда система находится в состоянии $|\psi\rangle$, среднее значение наблюдаемой w считается равным

$$\langle w\rangle = \langle\psi|W|\psi\rangle, \tag{5.1}$$

где W — эрмитов оператор, сопоставленный w. Операторы, сопоставленные наблюдаемым, удовлетворяют уравнениям движения и коммутационным соотношениям, которые иногда могут быть выведены из *правила коммутации*. Это правило гласит, что некоторые классические скобки Пуассона

$$\{a, b\} \equiv \sum_A \frac{\partial a}{\partial q_A}\frac{\partial b}{\partial p^A} - \frac{\partial a}{\partial p^A}\frac{\partial b}{\partial q_A} \tag{5.2}$$

переходят в коммутаторы, согласно соотношению

$$[A, B] \equiv AB - BA = i\hbar\{a, b\}. \tag{5.3}$$

Если скобки Пуассона $\{a, b\}$ в соотношении (5.3) равны константе, то они считаются оператором, являющимся c-числом. Другими словами, это единичный оператор из $\mathscr{H}$, умноженный на константу. Если скобки $\{a, b\}$ зависят от канонических переменных, то они должны быть заменены соответствующим оператором. Эта замена проводится непосредственно только тогда, когда скобки $\{a, b\}$ линейны по каноническим переменным (т. е. когда канонические переменные образуют группу Ли).

Для некоторых теорий, в частности для ОТО, правило коммутации должно быть изменено. Чтобы понять необходимые изменения, полезно вывести это правило таким путем, который выделяет его существенный физический смысл из изменчивого формального содержания. Делаются следующие физические предположения:

P1. Существуют *предпочтительные наблюдаемые*, такие, что если a и b принадлежат к ним, то канонические преобразования

$$a' = a + \varepsilon\{a, b\} + O(\varepsilon^2) \tag{5.4}$$

отождествляются с унитарными преобразованиями

$$A' = \left[\exp\left\{i\left(\frac{\varepsilon}{\hbar}\right)K(b)\right\}\right] A \left[\exp\left\{-i\left(\frac{\varepsilon}{\hbar}\right)K(b)\right\}\right] \tag{5.5}$$

соответствующих операторов в пространстве $\mathscr{H}$. Здесь $K(b)$ — *квантовый генератор*, соответствующий *классическому генератору* b, который еще должен быть определен и не обязательно является квантовым оператором B, соответствующим *наблюдаемой* b.

P2. Квантовые средние значения (5.1) предпочтительных наблюдаемых удовлетворяют классическим уравнениям движения.

Из P1, используя уравнения (5.1), (5.4) и (5.5), можно вывести уравнение

$$\left(\frac{\varepsilon}{\hbar}\right)\langle\psi|[A, K(b)]|\psi\rangle = i\varepsilon\{a, b\} + i(\langle\psi|A'|\psi\rangle - a') - i(\langle\psi|A|\psi\rangle - a)$$

для любого физического состояния $|\psi\rangle$. Из P2 следует заключение, что a может быть положено равным $\langle a\rangle$, а a' — равным $\langle a'\rangle$, так что для любого физического состояния

$$\langle\psi|[A, K(b)]|\psi\rangle = i\hbar\{a, b\}. \tag{5.6}$$

Поскольку эти рассуждения можно повторить и для переставленных a и b, находим также, что

$$\langle\psi|[A, K(b)] - [K(a), B]|\psi\rangle = 0 \tag{5.7}$$

для любого физического состояния $|\psi\rangle$. Этот последний результат является существенным звеном, связывающим квантовый генератор $K(b)$ с квантовой наблюдаемой B.

Чтобы получить обычную форму (5.3) правила коммутации, необходимо сделать дальнейшие предположения. Прямой подход, который удовлетворителен для простых систем, содержит формальное допущение (F0), что все состояния в гильбертовом пространстве $\mathcal{H}$ физически реализуемы. Тогда уравнение (5.7) имеет решения

$$K(b) = B + cI + d_s J_s, \tag{5.8}$$

где I — единичный оператор, а J_s ($s = 1, 2, 3, \ldots$) — полный набор операторов, коммутирующих со всеми предпочтительными наблюдаемыми. Этот результат вместе с уравнением (5.6) и эрмитовостью A и B приводит непосредственно к правилу (5.3). Для теорий, которые усложнены такими формальными требованиями, как явная ковариантность и калибровочная инвариантность, необходимо выбрать менее прямой подход, в котором $\mathcal{H}$ содержит *нефизические состояния*. Эти состояния не реализуются ни в каком эксперименте, и уравнения (5.6) и (5.7) не обязательно справедливы для них. Налагая формальное допущение (F1), что операторы A, B, $K(a)$ и $K(b)$ не содержат матричных элементов, связывающих физические и нефизические состояния, можно получить промежуточную форму правила коммутации

$$K(b)|\psi\rangle = (B + cI + d_s J_s)|\psi\rangle, \tag{5.9}$$

$$[A, B]|\psi\rangle = \hbar\{a, b\}|\psi\rangle \quad (\text{положив } \langle\psi|\psi\rangle = 1), \tag{5.10}$$

где $|\psi\rangle$ — физическое состояние. Чтобы получить сильную форму этого правила [уравнения (5.8) и (5.3)], требуется еще одно формальное предположение (F2) о том, что A и B дают нуль, когда они действуют на нефизические состояния.

Получение правила коммутации (5.3), (5.8), так же как его более слабых форм (5.6), (5.7) и (5.9), (5.10), оказывается полезным при обсуждении квантовой ОТО. Прежде всего из уравнений (5.8) и (5.9) ясно, что генератор $K(b)$ не обязательно должен быть равным соответствующей наблюдаемой B. Использование нормального упорядочения в квантовой теории поля является одним из примеров такой свободы при выборе генератора. Далее можно получить правило коммутации даже тогда, когда множество предпочтительных наблюдаемых не является полным коммутирующим набором. Однако квантовые генераторы, а для этого случая и все квантовые операторы, определяются затем лишь с точностью до произвольного члена. Этот пример важен в том случае, когда пытаются описать бесконечномерную теорию (например, ОТО) как предел последовательности конечномерных теорий (см. § 7, п. 4). В этом случае мы *никогда* не будем иметь полного набора предпочтительных наблюдаемых. Хуже того, результирующая неоднозначность каждой урезанной теории в пределе не стремится, вообще говоря, к нулю. Различные методы приближения могут привести к унитарно неэквивалентному набору квантовых операторов (см. [75], в особенности части 6 и 7). Наконец, допущение P2 о том, что предпочтительные наблюдаемые должны иметь строго классические средние значения, ясно предупреждает, что предпочтительные наблюдаемые следует выбирать с большой осторожностью и что правило коммутации не может выполняться для всех наблюдаемых.

Непосредственное доказательство необходимости ограничить правило коммутации (5.3) небольшим количеством предпочтительных наблюдаемых предложил Комар (неопубликованный конспект лекций, 1968 г.). Рассмотрим частицу в одном измерении. Используем координатное представление $X = x$, $P = -i\hbar d/dx$ для вычисления коммутатора $[\sin(2\pi X/l), \sin(lP)] = 0$ [1]). Теперь вычислим соответствующие скобки Пуассона

$$\left\{\sin\left(\frac{2\pi x}{l}\right),\ \sin(lp)\right\} = 2\cos\left(\frac{2\pi x}{l}\right)\cos(lp) \neq 0.$$

В канонической схеме квантования необходимо сначала решить, когда применять правило коммутации, а затем доказать, что

[1]) Будем рассматривать $\sin(lP)$ как конечноразностный оператор. Примером совместного собственного состояния коммутирующих операторов $\sin(lP)$ и $\sin(2\pi X/l)$ является состояние первоначально плоской волны сразу после того, как она прошла через бесконечную решетку с периодом l.

результирующая схема по крайней мере самосогласована. В примере Комара применение этого правила к x, p, $\sin(2\pi x/l)$ и $\sin(lp)$ не является согласованным.

В квантовой ОТО не было явного отождествления предпочтительных наблюдаемых. Однако, по крайней мере в виде пробы, Уилер [92] и де Витт [24] предложили в качестве предпочтительных наблюдаемых взять канонический гамильтониан (4.5), пространственные компоненты метрического тензора g_{ij} и сопряженные им импульсы π^{ij}, а также функции смещения и сдвига N_μ с сопряженными им импульсами π^μ. Они предложили представлять квантовое состояние функционалом $\Psi[g_{ij}, N_\mu]$ и определили операторы, действующие на такое состояние, посредством соотношений

$$g_{ij}(P)|\Psi\rangle = |g_{ij}(P)\Psi\rangle, \tag{5.11}$$

$$N_\mu(P)|\Psi\rangle = |N_\mu(P)\Psi\rangle, \tag{5.12}$$

$$\pi^{ij}(P)|\Psi\rangle = -iL^2\left|\frac{\delta\Psi}{\delta g_{ij}(P)}\right\rangle, \tag{5.13}$$

$$\pi^\mu(P)|\Psi\rangle = -iL^2\left|\frac{\delta\Psi}{\delta N_\mu(P)}\right\rangle \tag{5.14}$$

по аналогии с координатным представлением в обычной квантовой механике[1]). Здесь L — постоянная длина, которая появляется вследствие того, что правило коммутации применимо к импульсам, полученным из динамического действия

$$\frac{c^3}{16\pi k}\int d^4x\,(g^{1/2}R),$$

а не к импульсам π^{ij} и π^μ, которые получены из геометрического действия. Эта длина $L=(8hk/c^3)^{1/2}\approx 10^{-33}$ см подробнее рассматривается в п. 3 данного параграфа.

Поскольку π^μ является предпочтительной наблюдаемой, ее среднее значение удовлетворяет классическому уравнению

$$\langle\pi^\mu\rangle = \langle\Psi|\pi^\mu|\Psi\rangle = 0 \tag{5.15}$$

---

[1]) Здесь P обозначает точку в пространстве, а $\delta\Psi/\delta g_{ij}(P)$ является дифференциальным функционалом, определенным как

$$\int d^3x\left\{\frac{\delta\Psi}{\delta g_{ij}}v_{ij}\right\} \equiv \frac{\partial}{\partial\varepsilon}\Psi[g_{ij}+\varepsilon v_{ij}]\Big|_{\varepsilon=0}$$

для v_{ij} произвольных везде, за исключением граничных условий, позволяющих пренебречь интегралом по поверхности.

для всех физических состояний $|\Psi\rangle$. Положим в уравнении (5.6) $K(b)$ равным общековариантному гамильтониану $H = L^{-2}\int \sigma N_\mu R^\mu$, а $A = \pi^\mu$. Уравнение (5.6) тогда примет вид

$$\langle\Psi\,|[\pi^\mu,\,H]|\,\Psi\rangle = i\{\pi^\mu,\,H\} = -\,i\,\frac{\partial\pi^\mu}{\partial t} = 0,$$

или, так как $[\pi^\mu, H] = R^\mu$ [используем уравнение (5.14)],

$$\langle\Psi|R^\mu|\Psi\rangle = 0 \tag{5.16}$$

для всех физических состояний. На этом этапе не сделано каких-либо предположений о нефизических состояниях. В теории Уилера — де Витта эти предположения не могут быть сделаны произвольно, так как они уже неявно содержатся в уравнениях (5.11) — (5.14). Допущение F0 или комбинация допущений F1+F2 привели бы к результатам $\pi^\mu \equiv 0$, $R^\mu \equiv 0$ и $H \equiv 0$, которые явно опровергаются очевидно ненулевыми выражениями, следующими из уравнений (5.11) — (5.14). Единственным возможным предположением, которое может быть сделано относительно нефизических матричных элементов π^μ, R^μ и H, является F1, так что уравнения (5.15) и (5.16) могут быть усилены только до

$$\pi^\mu|\Psi\rangle = 0, \tag{5.17}$$

$$R^\mu|\Psi\rangle = 0 \tag{5.18}$$

для всех физических состояний $|\Psi\rangle$. Эти уравнения были предложены ранее Дираком [29].

Квантовая ОТО, предложенная Уилером и де Виттом, подобна по своей структуре фермиевской формулировке электродинамики [34]. Ферми рассматривает электрический потенциал (аналогичный N_μ) и продольную составляющую электромагнитного поля (аналогичную «зависимой от координат части g_{ij}») в качестве предпочтительных наблюдаемых и приходит к условиям для функционала состояния, которые аналогичны уравнениям (5.17) и (5.18). Однако Ферми мог *решить все полученные уравнения связи*. Он нашел, что функционал электромагнитного состояния является произведением особого функционала, зависящего от электрического потенциала и продольной компоненты, и функционала Ф, зависящего только от поперечных компонент. Более того, когда он подставил полный функционал состояния в основанное на полном каноническом гамильтониане уравнение Шредингера, то получил вполне разумное уравнение Шредингера для Ф.

Все общерелятивистские уравнения связи (5.17) и (5.18) могут быть легко решены, кроме одного. Четыре квантовых первичных уравнения связи (5.17) означают лишь, что физические функционалы состояния не зависят от функций смещения и сдвига N_μ. Три пространственноподобных вторичных уравнения связи [(5.18) для $\mu = 1, 2, 3$] при применении данного в уравнении (3.19) определения R^i и уравнения (5.13) сводятся к уравнениям

$$\left(\frac{\partial \Psi}{\partial g_{ij}}\right)_{\|j} = 0,$$

которые удовлетворяются, если Ψ является независимым от координат функционалом внутренней геометрии, описываемой посредством g_{ij} (см. [3, 50]). Единственное остающееся уравнение связи

$$R^0 |\Psi({}^3\mathcal{G})\rangle = 0$$

называется *гамильтоновой связью*, так как оно эквивалентно уравнению

$$\mathcal{H}(P) |\Psi({}^3\mathcal{G})\rangle = 0, \tag{5.19}$$

где $\mathcal{H}(P)$ — плотность гамильтониана $\dot{N}_\mu \pi^\mu + N_\mu R^\mu$ общековариантной канонической теории [1]). Это то самое уравнение связи, которое отличает квантовую ОТО Уилера — де Витта от электродинамики Ферми. Уравнение Шредингера, которое получается из канонического гамильтониана, устанавливает лишь, что $\Psi({}^3\mathcal{G})$ не зависит от какого-либо временного параметра.

Уравнение (5.19) и следующее из него простейшее уравнение Шредингера имеют простое объяснение. Неверно думать, что $\Psi({}^3\mathcal{G})$ — функционал внутренней геометрии пространства в «выделенный момент времени», поскольку, как показали Байерлейн и др. [7], внутренняя геометрия сама содержит всю динамически рассматриваемую информацию о времени.

В этой общековариантной квантовой теории функционал состояния Ψ не может явно зависеть от выбора временной координаты. Это требование означает, что Ψ не может зависеть от какого-либо временного параметра. Динамику можно ввести лишь одним способом — определить физическое время, которое является функционалом трехмерной геометрии, и решить затем гамильтоновы уравнения связи (которые заданы в каждой пространственной точке).

---

[1]) Здесь операторы π^μ и R^μ должны находиться справа от операторов $\dot{N}_\mu$ и N_μ.

Де Витт записывает гамильтоновы уравнения связи в форме

$$\left\{G_{ijkl}\frac{\delta}{\delta g_{ij}}\frac{\delta}{\delta g_{kl}}+g^{1/2}R\right\}\Psi\,[{}^3\mathscr{G}]=0, \tag{5.20}$$

где

$$G_{ijkl}=\frac{1}{2}g^{-1/2}(g_{ik}g_{jl}+g_{il}g_{jk}-g_{ij}g_{kl}). \tag{5.21}$$

Чтобы получить эту форму уравнений, он решает проблему упорядочения сомножителей, предполагая, что в данной пространственной точке коммутируют любые операторы [24]. Дельта-функции Дирака, которые входят в канонические коммутационные соотношения, строятся так, что $\delta(0)=0$. Вторые вариационные производные в уравнении (5.20) определяются тождеством

$$\left\{\frac{\delta}{\delta\phi(P)}\right\}^2 F\equiv\lim_{P\to P'}\left\{\frac{\delta^2 F}{\delta\phi(P)\,\delta\phi(P')}\right\},$$

где $\delta_2 F/\delta\phi(P)\,\delta\phi(P')$ имеет обычное определение

$$\int d^3x\int d^3x'\left\{\frac{\delta^2 F}{\delta\phi(P)\,\delta\phi(P')}\right\}v(P)\,v(P')\equiv\left.\frac{\partial^2 F(\phi+\varepsilon v)}{\partial\varepsilon^2}\right|_{\varepsilon=0}.$$

Герлах [41] показал, что вся классическая ОТО (т. е. десять уравнений Эйнштейна) может быть получена из одного уравнения связи (5.19). Герлах использует методы ВКБ, чтобы исследовать классический предел, в котором наиболее вероятные пространственные геометрии связаны вместе как пространственные сечения одного пространственно-временного многообразия. Правдоподобным аргументом в пользу этого результата является тот факт, что гамильтоново уравнение связи влечет за собой $\langle G_0^0\rangle=0$, не давая при этом никакой информации о том, как получить в классическом пространстве-времени компоненту G_0^0 4-тензора G_λ^μ. При условии, что $\langle G_\nu^\mu\rangle$ может быть отождествлено с тензором в классическом пространстве-времени, этот факт приводит к тому, что $\langle G_\nu^\mu\rangle$ обращается в нуль для всех μ и ν.

Как мы увидим ниже при обсуждении квантованных замкнутых моделей вселенной (§ 7), подход Уилера — де Витта еще далек от полностью разработанной квантовой теории ОТО вследствие наличия нерешенных проблем интерпретации, таких, как определение скалярных произведений между состояниями. Далее мы сталкиваемся с произволом выбора g_{ij} в качестве предпочтительных координат, которые должны удовлетворять обычным коммутационным соотношениям. Клаудер [53] показал, что положительно определенная сигнатура g_{ij} исключает любые самосопряженные представления π^{ij} аналогично тому, как положительная определенность энергии в обычной квантовой механике исключает самосопряженное выражение оператора времени. Если $P=k$ положительно определено и каноническая сопряженная величина $Q=-i(\partial/\partial k)$ предполагается самосопряженной, то можно использовать унитарные операторы $U(a)=\exp(iaQ)$, где a может быть любым действительным числом, чтобы получить отрицательные собственные значения P. И наоборот, можно показать, что операторы $U(a)$ не могут быть унитарными, так как обратную величину нельзя определить в пространстве положительных собственных значений P.

Один из способов рассмотрения проблемы положительной определенности состоит в том, чтобы пренебречь, как в теории Уилера — де Витта, отсутствием самосопряженных представлений импульсов. Тогда можно быть вполне уверенным в том, что измеримые комбинации g_{ij} и π^{ij} являются самосопряженными. Другой способ рассмотрения этой проблемы состоит в том, чтобы определить новые переменные $\beta_{ij}(P)$ таким образом, что

$$\{\exp\beta(P)\}_{ij}\equiv\delta_{ij}+\beta_{ij}(P)+\frac{1}{2}\sum_k\beta_{ik}\beta_{kj}+\ \ldots\equiv g_{ij}(P).$$

Этот «экспоненциальный путь» был избран Мизнером в его анализе квантовой космологии (§ 7, п. 4). Трудность здесь заключается в том, что функционалы лагранжиана и гамильтониана, выраженные через экспоненциальные переменные, не обладают простой структурой. Клаудер предлагает еще один путь, который сохраняет и алгебраическую форму гамильтониана, и самосопряженность основных квантовых операторов (см. также [6]). Он определяет самосопряженное выражение оператора B, которое удовлетворяет коммутационному соотношению $[P, B]=$ $=iP$. Это соотношение отождествляет P и B как генераторы аффинных преобразований действительной прямой. Существует представление аффинной группы, для которого P положительно определено [56]. «Аффинно сопряженный» с P оператор B связан с канонически сопряженным оператором Q соотношением $B=$ $={}^1/_2(QP+PQ)$. Поскольку операторы, используемые в этих трех подходах к проблеме положительной определенности, локально связаны друг с другом, в принципе можно перевести квантовую ОТО с языка одного подхода на язык другого подхода. Однако, как указал Клаудер, кажущийся естественным в одном подходе анализ таких неоднозначностей, как упорядочение сомножителей и представление коммутационных соотношений, может оказаться совершенно неестественным при переводе на язык другого подхода.

Теория Уилера — де Витта достигает явной независимости от координат за счет применения очень необычного формализма, в котором время теряет свой обычный смысл, а операторы определяются в гильбертовом пространстве, дополненном в значительной степени нефизическими состояниями. Введение нефизических состояний для достижения явной ковариантности или калибровочной инвариантности не является необычным в квантовых теориях поля, но теория Уилера — де Витта доводит этот процесс до такой степени, что уравнения связи, определяющие физическое подпространство, приобретают смысл всей теории.

Жертвуя явной ковариантностью, можно достичь более привычного вида канонического квантования ОТО. По существу в качестве модели для квантования берется обычный канонический формализм вместо общековариантного канонического формализма. Одна из возможностей состоит в том, чтобы решить уравнения связи и координатные условия до квантования. Полевые переменные в такой теории описывают только физические частицы. Квантование проводится непосредственно и свободно от калибровочных трудностей и трудностей, связанных с наличием уравнений связи. Поскольку уравнения связи определены только по отношению к данным пространственноподобным гиперповерхностям, этот подход нарушает явную лоренц-ковариантность и таким образом заставляет проверять лоренц-ковариантность окончательных амплитуд рассеяния. Этот нековариантный подход использует Мизнер (§ 7, п. 4). Другой возможностью является наложение после квантования дополнительных лоренц-ковариантных условий и решение уравнений связи как квантовых (q-числовых) уравнений. Получаемая теория аналогична формализму Гупта — Блейера в электромагнитной теории. Лоренц-ковариантность получается за счет введения состояний с отрицательными нормами. Этому подходу следуют Гупта, Баркер и др. [8, 9, 47]. Более подробное рассмотрение обычной формы квантовой ОТО мы отложим до § 6, где она естественно появляется при квантовании открытых моделей вселенной.

2. *Сумма по историям: процедура вычислений*

Фейнмановская процедура суммирования по историям представляет собой метод квантования, который пытается обойти канонический формализм. Фейнмановские правила построения амплитуд вероятности в том виде, как они обычно применяются, не позволяют получить основу теории квантования. Скорее они обеспечивают форму, к которой всегда могут быть приспособлены квантовые вычисления, независимо от их конкретной структуры [35, 36].

Основная величина, которую вычисляет квантовая теория,— это вероятность получить тот или иной результат в конкретном эксперименте. Квантовые эксперименты всегда оставляют неопределенным (с некоторыми ограничениями) «путь» или «физический процесс», посредством которого получается данный результат. Законом природы, полученным из многих экспериментов, является тот факт, что альтернативные процессы являются либо интерферирующими, либо неинтерферирующими. Каждому процессу сопоставляется комплексное число $A(b)$, называемое амплитудой вероятности. Смысл этих чисел, а также терминов «интерферирующий» и «неинтерферирующий» устанавливается первым правилом Фейнмана: если результат c может появиться в результате двух процессов a и b, то вероятность получить c равна

$$P(c)=\begin{cases}\{A(a)+A(b)\}\{A(a)+A(b)\}^{*}, & \text{если } a \text{ и } b \text{ интерферируют},\\ A(a)A^{*}(a)+A(b)A^{*}(b), & \text{если } a \text{ и } b \text{ не интерферируют}\end{cases}$$

(звездочка обозначает комплексное сопряжение). Процессы, приводящие к различным экспериментальным результатам (т. е. регистрируемым данным), никогда не интерферируют. Процессы, приводящие к общему результату (промежуточные процессы), могут как интерферировать, так и не интерферировать. В этих рамках только эксперимент может определить, какой случай имеет место.

Второе правило говорит о том, что амплитуды для процессов, происходящих в последовательные промежутки времени, должны умножаться. Если $\Psi_{k,t_i}(x_i)$ — амплитуда того, что частица в состоянии k имеет координату x_i в момент t_i, а $K_{tt'}(x, x')$ — амплитуда того, что частица имеет координату x в момент t и координату x' в момент t', то амплитуда вероятности того, что частица имеет последовательность пространственно-временных координат (x_i, t_i), (x, t), (x_f, t_f), есть

$$A(x_i, t_i;\ x, t;\ x_f, t_f)=K_{t_it}K_{tt_f}.$$

Амплитуда нахождения частицы в состоянии k, так же как и ее пространственно-временных локализаций (x_i, t_i) и (x_f, t_f), есть

$$A(k;\ x_i, t_i;\ x_f, t_f)=\Psi_{k,t_i}K_{t_it_f}.$$

Амплитуда того, что частица будет двигаться по «скелетной» траектории через последовательность положений

$$(x_i,\ t_i),\ (x_1,\ t_1),\ (x_2,\ t_2),\ \ldots,\ (x_n,\ t_n),\ (x_f,\ t_f),$$

есть «скелетная амплитуда»

$$A(x_i,\ t_i;\ x_1,\ t_1;\ x_2, t_2;\ \ldots;\ x_n, t_n;\ x_f, t_f) = K_{t_i t_1} K_{t_1 t_2} K_{t_2 t_3} \ldots K_{t_n t_f}.$$

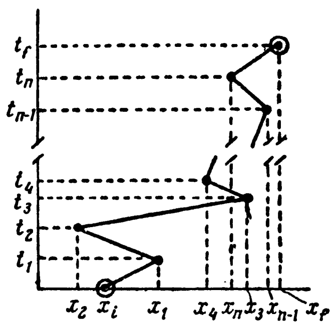
На основании экспериментального результата, что различные скелетные пути интерферируют между собой (двухщелевой эксперимент), можно записать следующие соотношения для амплитуд:

$$\Psi_{k,\,t_f}(x_f) = \int dx_i K_{t_i t_f}(x_i,\ x_f)\,\Psi_{k t_i}(x_i),$$

$$K_{t_i t_f} = \int dx K_{t_i t} K_{t t_f}, \tag{5.22}$$

$$K_{t_i t_f} = \int dx_1 \int dx_2 \ldots \int dx_n K_{t_i t_1} K_{t_1 t_2} \ldots K_{t_n t_f}. \tag{5.23}$$

Таким образом, мы видим, что $K_{t_i t_f}$ — пропагатор, или функция Грина, которая описывает развитие состояния во времени.



Фиг. 2. Каждая линия представляет пропагатор; пропагатор от i до f получается суммированием скелетных историй по всем $x_1,\ x_2, \ldots, x_{n-1}$ и x_n.

Пространственно-временные положения, которые суммируются в выражении (5.23), показаны на фиг. 2.

Выражение (5.23) по существу представляет собой выражение для фейнмановского суммирования по историям для пропагатора. Чтобы привести его к более привычному виду, определим среднее положение $x_{12} \equiv {}^1/_2\,(x_1 + x_2)$, среднюю скорость $v_{12} \equiv (x_2 - x_1)/(t_2 - t_1)$ и выразим пропагатор в совершенно общем виде

$$K_{t_1 t_2}(x_1,\ x_2) \equiv$$
$$\equiv N_{t_1 t_2}(x_1,\ x_2)\exp\left\{\frac{i}{\hbar}(t_2 - t_1)\,L(v_{12},\ x_{12},\ t_1,\ t_2)\right\}. \tag{5.24}$$

где N и L — вещественные функции и $\hbar$ — постоянная Планка $(h/2\pi)$. Перепишем теперь выражение (5.23) в виде

$$K_{t_i t_f} = \lim_{m\to 0} \int dx_1 \ldots \int dx_n N_{t_i t_1} \ldots N_{t_n t_f} \exp\left\{ \frac{i}{\hbar} \int_i^f dtL \right\}, \quad (5.25)$$

где $\int_i^f dt\,L$ — лишь сокращенное обозначение суммы

$$\sum_i^f (t_{j+1} - t_j)\, L\,(v_{j+1,\,j},\ x_{j+1,\,j},\ t_{j+1},\ t), \quad (5.26)$$

а m — порядок разбиения, или наибольшая из разностей $t_{j+1} - t_j$. Полезность выражения (5.25) следует из того обстоятельства, что при m, стремящемся к нулю, сумма (5.26) превращается в классический функционал действия, что и подразумевают наши обозначения. Легко показать (исходя из общепринятой квантовой теории), что это обстоятельство имеет место для систем Бозе — Эйнштейна, лагранжиан которых квадратичен по скоростям [39]. Клаудер [53] показал, что это применимо также для простых систем Ферми — Дирака, хотя для таких систем важность разрывных историй не позволяет получить правильное скелетное действие (5.26) из одного только классического действия. Стандартный путь дальнейшего сокращения выражения (5.26) следующий:

$$K_{t_i t_f} = \int \mathscr{D}x \exp\left\{ i\frac{I}{\hbar} \right\}, \quad (5.27)$$

где I — действие $\int dt\,L$.

Чтобы увидеть, как применяются правила Фейнмана в общепринятой квантовой теории, введем обозначения

$$\langle x,\ t\,|\,k\rangle = \Psi_{kt}(x)$$

и

$$\langle x_f,\ t_f\,|\,x_i,\ t_i\rangle = K_{t_i t_f}(x_i,\ x_f),$$

чтобы рассматривать сумму по историям как инструмент для получения конечного преобразования представления (от базиса $|x_i, t_i\rangle$ к базису $|x_f, t_f\rangle$) путем подсчета инфинитезимальных преобразований базиса [1]).

[1]) Это представление Гейзенберга, в котором $|x, t\rangle$ — собственный вектор $X(t)$. В картине Шредингера сумма по историям является решением зависящего от времени уравнения для вектора состояния $|\,k\rangle_t = \int dx \Psi_{kt}(x)\,|\,x\rangle$.

Основная трудность, связанная с выражениями для сумм по историям (5.25)—(5.27) для пропагатора,— это хорошо известная проблема выбора меры. Необходимо решить, какими должны быть веса $N_{t_i t_1}$, $N_{t_1 t_2}$, ... Согласно критерию, использованному Лейтвилером [63] в его работе по квантованию ОТО на основе суммирования по историям, следует положить веса равными

$$N_{tt'}(x, x') = n(t, t') f(x) f(x').$$

В этом случае достаточно потребовать, чтобы пропагатор подчинялся условию

$$\lim_{t' \to t} K_{tt'}(x, x') = \delta(x - x'),$$

а также удовлетворял правилу композиции (5.22). Используя ковариантные компоненты метрического тензора $g_{\mu\nu}(P)$ как переменные интегрирования, Лейтвилер нашел, что весовой вклад для пространственно-временных историй равен

$$f(g_{\mu\nu} : \Sigma) = \text{const} \times \Pi\ |^4 g|^{-5/2} |\,^3 g\,|, \tag{5.28}$$

где $g : \Sigma$ представляет функции $g_{\mu\nu}(P)$, определенные всюду на пространственноподобной гиперповерхности Σ при постоянном x^0, а Π обозначает «произведение» по всем точкам Σ. Другой критерий выбора меры, предложенный Мизнером [66], подчеркивает требование, что операция суммирования по историям должна быть инвариантной относительно группы «однородных преобразований», которая явно локальна (нет ни пространственных, ни временных производных, а также интегралов) и включает в себя симметрию и калибровочные преобразования классической теории. Для ОТО он доказал, что соответствующими преобразованиями являются $g_{\mu\nu} \to g_{\mu\nu}^{(m)}$, где

$$g_{\mu\nu}^{(m)} = m_\mu^\alpha g_{\alpha\beta} m_\nu^\beta,$$

а m — произвольные пространственно-временные функции. Эти преобразования порождают весовые вклады сумм по историям, эквивалентные

$$f(g : \Sigma) = \text{const} \times \Pi\, |^4 g|^{-5/2}. \tag{5.29}$$

Результат Мизнера для весового вклада был получен также из несколько иных соображений Лаурентом [61]. Последний указал, что если действие для ОТО записать в виде

$$I_{\Gamma\Gamma} = \int {}^4\sigma G^{\alpha\beta\mu\nu\sigma\tau} g_{\alpha\beta,\sigma} g_{\mu\nu,\tau},$$

то весовой вклад Мизнера принимает вид

$$f(g:\Sigma) = \mathrm{const} \times \Pi \left\{ \det g_{\sigma\tau} G^{(\alpha\beta)(\mu\nu)\sigma\tau} \right\}^{1/2}, \qquad (5.30)$$

где определитель берется от матрицы (10×10) с индексами $(\alpha\beta)$ и $(\mu\nu)$. Это доказывает, что весовой вклад Мизнера эквивалентен взятию $G^{(\alpha\beta)(\mu\nu)\sigma}{}_{\sigma}\delta(P, P')$ в качестве компонент метрического тензора на множестве всех $g_{\mu\nu}$.

Мизнер получил свою группу однородности расширением группы преобразований координат. Де Витт [21] отметил, что любое такое расширение довольно искусственно, что группа преобразований координат одна не может определять относительные весовые вклады для определенной геометрии пространства-времени. Де Витт рассматривает неопределенную часть весового вклада как средство для образования контрчленов, которые обеспечивают эрмитовость уравнений поля и унитарность S-матрицы [26].

Критерии Мизнера и Лейтвилера имеют довольно ограниченное значение, так как исключена зависимость весового вклада от производных по времени. Например, эти критерии могут быть использованы для релятивистской частицы с времениподобной траекторией лишь в том случае, если последняя достаточно удобна для работы в импульсном представлении. В этом представлении группа однородности Мизнера легко может быть получена ослаблением дифференциальных ограничений

$$\frac{\partial m^{\mu}_{\nu}}{\partial x^{p}} = 0$$

на преобразования Лоренца для p^{μ}:

$$p^{\mu}(m) = m^{\mu}_{\nu} p^{\nu}.$$

Мера в пространстве импульсов d^3p/p^0 удовлетворяет предположению Лейтвилера, так как она не содержит производных по времени от p^1, p^2 и p^3. В координатном пространстве невозможно строго локальное расширение группы инвариантности Пуанкаре без нарушения условия, что траектории должны оставаться внутри светового конуса. Более того, хорошо известная форма $\int d^3x \psi^* \overleftrightarrow{\partial_0} \psi$ релятивистского скалярного произведения в координатном пространстве может получаться только из зависящей от времени меры, которая нарушает основное предположение Лейтвилера.

Составной частью проблемы выбора меры, более тонкой, чем выбор веса $N_{tt'}$, является *аппроксимация функционалов-исто-*

рий. Чтобы определить сумму по историям, необходимо сначала определить ее для *простых интегрируемых функционалов*, например для функционалов, зависящих только от конечного числа переменных, которые определяют историю (каркас), или функционалов специального вида (в теории возмущений простыми интегрируемыми функционалами являются

$$F(n, n', \ldots, a, a', \ldots)[x(t)] \equiv x^n(a)\, x^{n'}(a') \ldots \exp[iI_0],$$

где I_0 — невозмущенный функционал действия). Чтобы проинтегрировать сложный функционал, нужно аппроксимировать его последовательностью простых интегрируемых функционалов. Тогда интеграл от сложного функционала *определяется* как предел последовательности простых интегралов. К сожалению, различные аппроксимирующие последовательности часто дают разные результаты для одного и того же интеграла. Например, не существует однозначного способа проинтегрировать функционал

$$F(t)[x(t)] \equiv x(t)\, \dot{x}(t).$$

В результате среднее значение $\langle x\dot{x}\rangle$ определяется неоднозначно. Неоднозначность результата в данном примере соответствует упорядочению некоммутирующих операторов X и P в общепринятой квантовой теории. В дополнение к таким проблемам упорядочения сомножителей выбор процедуры аппроксимации включает в себя также стандартную для теории поля проблему выбора канонических коммутационных соотношений среди многих унитарных неэквивалентных представлений (см., например, [75], стр. 388—394). Более того, Клаудер [53] показал, что различие между статистиками Бозе — Эйнштейна и Ферми — Дирака может рассматриваться, как появляющееся исключительно в результате выбора аппроксимирующих последовательностей для функционала действия.

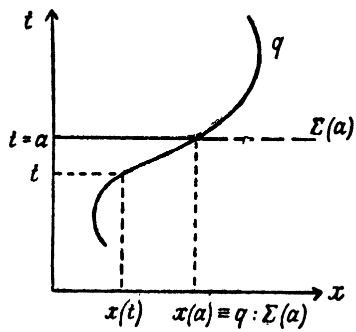
Процедура суммирования по историям использовалась Лейтвилером [63] для введения канонического формализма квантования, де Виттом [26] — для разложения гравитационной S-матрицы и Уилером — для обсуждения общей природы квантовой ОТО. Как для классического функционала действия редко ищут непосредственно его экстремум (вместо этого выводят уравнения Лагранжа), так и редко подсчитывают полную сумму по историям. Во всех приложениях, которые мы упомянули, суммирование по историям используется как средство для образования понятий и вывода более удобных теорий. Ниже будет показано, что абстрактный вариант суммирования по историям может быть использован для разработки многих стандартных процедур квантования.

3. *Векторы-истории: объединенный подход к квантованию*

Мы рассмотрели два различных подхода к квантованию: каноническое квантование в гильбертовом пространстве и применение правил Фейнмана. Правила Фейнмана не полны, но чрезвычайно гибки и трактуют все процессы с единой точки зрения. С другой стороны, каноническое квантование, по-видимому, представляет собой завершенную теорию, но с самого начала налагает на процессы жесткие ограничения, особенно на те из них, которые связаны с определением состояния системы.

В ОТО нам не хватает физической интуиции, которая могла бы дополнить правила Фейнмана (соответствующим выбором меры), а также твердого знания процедуры определения состояния. Именно по этим причинам некоторые авторы, например де Витт [26], используют оба подхода одновременно. Ниже мы предлагаем объединенный подход, который сочетает характерные черты обоих формализмов. В этом подходе используются элементы одного большого векторного пространства для физической нумерации состояний в любом возможном представлении квантового гильбертова пространства. Мы нашли, что такой объединенный подход весьма ценен для сопоставления различных формулировок квантовой ОТО.

Основное понятие, которое мы вводим, — это представление квантовых состояний в виде функционалов-историй. Это понятие может быть определено посредством набора абстрактных постулатов. Однако наиболее простой путь введения этих постулатов заключается в разборе несложного примера. Рассмотрим одномерный гармонический осциллятор. Обозначим смещение через x и время через t. Пусть $\Sigma(a)$ — совокупность всевозможных смещений в момент $t = a$. Назовем ее «равновременной гиперповерхностью». Обозначим траекторию, или историю, осциллятора через q и будем определять ее путем задания x как функции t. Как классическая физика, так и квантовое суммирование по историям выражаются в виде функционалов от таких историй (фиг. 3). Однако в общепринятой (гейзенберговской) квантовой теории состояние системы задается функцией состояния, аргументом которой является смещение частицы в выбранный момент времени $t = a$. Чтобы получить представление в виде функционала-истории, будем рассматривать $x(a)$ как смещение, которое определяется как пересечение истории q с равновременной гиперповерхностью $\Sigma(a)$. Чтобы показать, что смещение рассматривается как функционал по истории, достаточно записать его в виде $x(a) = q : \Sigma(a)$. Тогда для каждой функции состояния $\psi(x)$ можно определить соответствующий функционал-историю Ψ, потребовав, чтобы $\Psi[q] = \psi(q:\Sigma(a))$ для всех q. Простой способ, которым было определено соответствие, показывает, что линейные комбинации таких функционалов-историй представляют соответствующие линейные комбинации векторов состояния. Любое отображение функционалов-историй на векторы состояния, которое таким способом сохраняет линейную алгебру, мы будем называть *функциональным представлением* квантового гильбертова пространства. Если, как в нашем примере, отображение однозначно, то оно будет называться *изоморфным представлением*.



Фиг. 3. Смещение при $t = a$ может рассматриваться как функционал истории q; история q определяется функцией $x(t)$.

Могут быть полезны еще два примера представлений функционалов-историй одномерного гармонического осциллятора. Если принять, что $p(a) = q' : \Sigma(a)$ — классический импульс, определенный на истории q в момент времени $t = a$, то гейзенбергова функция состояния в импульсном пространстве $\psi(k)$ может быть представлена функционалом-историей Ψ, который определен как $\Psi[q] = \psi(q' : \Sigma(a))$. Эти функционалы-истории отличаются от тех, которые использовались в предыдущем примере, но они также определяют изоморфное представление гильбертова пространства. Здесь мы заменили квантовый импульс k классическим функционалом импульса $q' : \Sigma(a)$ в аргументе функции состояния. Так же легко мы могли бы заменить k классическим функционалом положения $q : \Sigma(a)$ и все еще получили бы изоморфное представление. Однако, используя операторные уравнения движения Гейзенберга для оценки средних значений x и p, можно быстро обнаружить, что q является историей в пространстве импульсов. Чтобы увидеть, как могут быть получены неизоморфные представления, рассмотрим дополненное множество состояний $|x, p\rangle$, например гауссовы минимально неопределенные волновые пакеты в момент времени $t=a$. Произвольное состояние может быть представлено в виде

$$|\psi\rangle=\int dx \int dp\psi(x,\ p)|x,\ p\rangle,$$

а соответствующий функционал-история определяется как

$$\Psi[q]=\psi(q:\Sigma(a),\ q':\Sigma(a)).$$

Эти функционалы образуют функциональное представление гильбертова пространства, но они не находятся во взаимно однозначном соответствии с отдельными квантовыми состояниями.

Из этих примеров можно вывести определяющие свойства представления в виде функционалов-историй. Во-первых, функциональное представление является множеством функционалов-историй вместе с отображением этого множества на гильбертово пространство квантовых состояний. Состояние, которое является образом функционала F, обозначается $|F\rangle$. Во-вторых, если строится *векторное пространство* $\mathscr{V}$ множества всех функционалов-историй путем задания таких линейных комбинаций $aF+bG$ функционалов F и G, что выполняется условие

$$\{aF+bG\}[q]=aF[q]+bG[q]$$

для всех историй q (a и b могут быть любыми комплексными числами), то соотношение

$$|aF+bG\rangle=a|F\rangle+b|G\rangle \tag{5.31}$$

должно выполняться для всех a и b, когда F и G принадлежат одному и тому же представлению. В-третьих, если F и G принадлежат представлению $\mathfrak{r}$, то это же имеет место для $aF+bG$,— представление $\mathfrak{r}$ является линейным подпространством векторного пространства $\mathscr{V}$. Поскольку функционалы-истории принадлежат векторному пространству, они могут быть названы *векторы-истории*. Чтобы завершить определение, мы должны определить *вариации-истории*. Вариация-история характеризуется для каждой истории q_0 однопараметрическим семейством историй $q[\varepsilon,q_0]$, таким, что $q[0,q_0]=q_0$, а ε является вещественным числом на отрезке между -1 и $+1$. Вариация δF функционала F в точке q_0 определяется как

$$\delta F[q_0]\equiv\frac{d}{d\varepsilon}F[q(\varepsilon,\ q_0)]|_{\varepsilon=0}\,d\varepsilon.$$

Для каждого функционального представления $\mathfrak{r}$ существует множество $\mathscr{C}_\mathfrak{r}$ вариаций $\delta_\mathfrak{r}$, таких, что

$$\delta_\mathfrak{r}F=0 \tag{5.32}$$

тогда и только тогда, когда F является функционалом в $\mathscr{r}$. Это характерное свойство представления $\mathscr{r}$ означает, что функционалы в $\mathscr{r}$ зависят только от той части информации, которая необходима для определения истории. Если $\mathscr{r}$ является рассмотренным выше координатным представлением, то все вариации в $\mathscr{C}_{\mathscr{r}}$ фиксируют $x(a) = q : \Sigma(a)$.

Чтобы произвести все квантовомеханические расчеты, используя одни только функционалы-истории, необходимо иметь выражения скалярного произведения для гильбертова пространства через функционалы-истории. Мы постулируем, что если имеется F в представлении $\mathscr{s}$, а G в представлении $\mathscr{r}$, то скалярное произведение $\langle F|G\rangle$ векторов состояния $|F\rangle$ и $|G\rangle$ есть

$$\langle F|G\rangle = \int_{\mathscr{r}}^{\mathscr{s}} \mathscr{D}q\,\{F^*G \exp(iI[\mathscr{r}, \mathscr{s}])\}, \tag{5.33}$$

где $I[\mathscr{r}, \mathscr{s}]$ — функционал-история, а $\int \mathscr{D}q \ldots$ — интегральный функционал — линейное отображение функционалов-историй на комплексные числа. Этот постулат является существенным обобщением фейнмановского интеграла по траектории или правила сумм по историям для вычисления матричных элементов пропагатора. Мы могли бы дополнить этот постулат, дав определение скелетного интеграла для $\int \mathscr{D}q$. Однако достаточно, а в некоторых случаях даже желательно принять более свободное определение, которое требует только следующих свойств:

1) вещественности произвольного функционала-истории Q, т. е.

$$\int_{\mathscr{r}}^{\mathscr{s}} \mathscr{D}q\,\{Q^*\} = \left(\int_{\mathscr{r}}^{\mathscr{s}} \mathscr{D}q Q\right)^*;$$

2) однородности произвольного функционала-истории, т. е.

$$\begin{gathered}\int_{\mathscr{r}}^{\mathscr{s}} \mathscr{D}q\,\{\delta_{\mathscr{r}} Q\} = 0 \quad \text{для любых} \quad \delta_{\mathscr{r}} \ \text{в} \ \mathscr{C}_{\mathscr{r}},\\ \int_{\mathscr{r}}^{\mathscr{s}} \mathscr{D}q\,\{\delta_{\mathscr{s}} Q\} = 0 \quad \text{для любых} \quad \delta_{\mathscr{s}} \ \text{в} \ \mathscr{C}_{\mathscr{s}}.\end{gathered} \tag{5.34}$$

Требование однородности может быть принято как теорема функционального «интегрирования по частям». Это требование ограничивает как множества вариаций, так и интегральный функционал. Множества $\mathscr{C}_r$ и $\mathscr{C}_s$ должны быть группами однородности, а $\int\limits_r^s \mathscr{D}q$ должен быть инвариантным относительно этих функциональных групп.

Здесь мы можем остановиться и построить такую форму квантовой теории, которая очень близка к фейнмановскому подходу сумм по историям. Однако полезно ввести представление функционалов-историй для операторов гильбертова пространства для установления связи с общепринятой квантовой теорией. Эти представления основаны на правиле функционального произведения:

$$FG[q] \equiv F[q]\,G[q] \quad \text{для всех историй} \quad q.$$

Конечно, произведение не сохраняет функциональных представлений так, как в случае линейной комбинации. Множество всех функционалов-историй вида FG, где F принадлежит r, а G принадлежит s, составляет *представление произведения* $r \times s$. Множество вариаций, которое определяет $r \times s$, есть $\mathscr{C}_{r \times s} = \mathscr{C}_r \cap \mathscr{C}_s$. Для данного функционала-истории Q и данного представления r определяют действие гильбертова оператора $T(r)[Q]$ на состояние $|F\rangle$, где F в представлении r, как

$$\{T(r)[Q]\}\,|F\rangle \equiv |QF\rangle. \tag{5.35}$$

Этот оператор отображает гильбертово пространство на себя, но он изменяет функциональное представление гильбертова пространства, если Q не в представлении r. Оператор $T(r)[Q]$ согласуется с общепринятым квантовым оператором, сопоставляемым наблюдаемой Q с точностью до обычного упорядочения сомножителей и перенормировочных неоднозначностей. В квантовой теории поля эти операторы T являются по существу теми операторами, которые появляются из хронологического упорядочения сомножителей.

Из постулата скалярного произведения совместно с требованием однородности можно получить тождество

$$\langle F\,|\{T(r)\,[\delta_{r \times s} I(r,\, s)]\}|\,G\rangle \equiv 0 \tag{5.36}$$

для всех F в представлении s и всех G в представлении r. Квантовый оператор, который соответствует $\delta_{r \times s} I$, имеет только нулевые матричные элементы. Таким образом, естественно, по крайней мере с точностью до обычных неоднозначностей, отождествить $I(\mathscr{r}, \mathscr{s})[q]$ с классическим действием истории q, а $\delta_{\mathscr{r}\times\mathscr{s}}$ — с вариацией такого типа, который дает классические уравнения движения без остаточных поверхностных членов в граничных точках. В той мере, в какой это отождествление справедливо, мы получаем не только функционал I, но и ценный ключ к природе вариации в $\mathscr{C}_{\mathscr{r}\times\mathscr{s}}$. С помощью этого ключа можно сделать вывод о природе фактор-множеств $\mathscr{C}_{\mathscr{r}}$ и $\mathscr{C}_{\mathscr{s}}$, которые определяют представления, связанные данным функционалом действия.

Из постулата скалярного произведения и однородности может быть получено также второе уравнение

$$\langle F|\{T(\mathscr{r})[\delta_{\mathscr{s}}I(\mathscr{r}, \mathscr{s})]\}|G\rangle \equiv -i\langle F|\delta_{\mathscr{s}}G\rangle, \qquad (5.37)$$

где F принадлежит $\mathscr{s}$, а G — $\mathscr{r}$. Этот результат полезен для установления связи между функциональным T-представлением гильбертовых операторов и представлениями в виде вариаций или частных производных.

Чтобы увидеть, как работает эта формальная техника, вернемся к гармоническому осциллятору. Предположим, что мы знаем только то, что классические уравнения движения могут быть получены варьированием функционала действия

$$\hbar I(\mathscr{f}, \mathscr{g})[q] = \int_a^b \frac{1}{2}(m\dot{x}^2 - kx^2)\,dt \qquad (5.38)$$

по функции $x(t)$. Поскольку функция $x(t)$, определенная на интервале $a \leqslant t \leqslant b$, должна варьироваться, мы заключаем, что эта функция полностью определяет историю q. Пока мы ничего не знаем о представлениях $\mathscr{f}$ и $\mathscr{g}$, которые связаны этим действием. Так как $x(a)$ и $x(b)$ должны быть фиксированы, чтобы избежать членов в граничных точках, мы по крайней мере знаем, что вариации $\delta_{\mathscr{f}\times\mathscr{g}}$ должны сохранять эти величины фиксированными. Мы заключаем, что $\mathscr{C}_{\mathscr{f}}$, которое характеризует представление $\mathscr{f}$, является набором вариаций, сохраняющих $x(a)$ фиксированной. Следовательно, $\mathscr{f}$ должно быть представлением положения при $t = a$. Точно так же $\mathscr{g}$ должно быть представлением положения при $t = b$.

Чтобы завершить квантовую теорию гармонического осциллятора, необходимо установить, какие из множества вариаций, сохраняющих $x(a)$ фиксированной, действительно принадлежат $\mathscr{C}_{\mathscr{f}}$. Простого способа получения этой информации из одного только функционала действия не существует, Корректная квантовая теория гармонического осциллятора получается сведением вариаций в $\mathscr{C}_{\mathscr{f}}$ к таким вариациям, для которых $\delta_{\mathscr{f}} x(t)$ не зависит от истории q_0, на которой она определяется. Такой выбор «группы однородности» соответствует вычислению сумм по историям путем интегрирования по $x(t)$ в каждый момент времени t с весовым множителем, равным единице. Приняв

$$F[q] = \delta(q : \Sigma(b) - x'), \quad G[q] = \delta(q : \Sigma(a) - x),$$

$\mathscr{r} = \mathscr{f}$ и $\mathscr{s} = \mathscr{g}$ в определении скалярного произведения, мы можем получить фейнмановские правила сумм по историям для оценки пропагатора $K_{ab} = \langle x', b | x, a \rangle$.

Рассмотрим теперь другой функционал действия для гармонического осциллятора. Действие

$$I(\mathscr{f}, \mathscr{p}) \equiv I(\mathscr{f}, \mathscr{g}) - x(b) p(b) \tag{5.39}$$

требует вариаций, которые фиксируют $x(a)$ и $p(b)$, чтобы избежать членов в граничных точках. Очевидно, $\mathscr{p}$ является импульсным представлением. Используя постулат о скалярном произведении, получаем снова правила Фейнмана с автоматически появляющимся требуемым множителем $\exp(-ixp)$. Когда b стремится к a, мы получаем обычные трансформационные соотношения Фурье между координатными и импульсными представлениями.

В качестве последнего примера из теории функционала-истории гармонического осциллятора рассмотрим уравнение (5.37) для $\mathscr{r} = \mathscr{f}$ и $\mathscr{s} = \mathscr{g}$. После использования уравнения (5.36) для устранения члена движения левая часть уравнения (5.37) будет содержать представление T для оператора импульса, а правая часть — обычное представление в частных производных. Следовательно, эти представления эквивалентны, хотя форма представления в частных производных сохраняет функциональное представление, в то время как T-форма переводит вектор из представления $\mathscr{f}$ в представление $\mathscr{f} \times \mathscr{p}$.

Обозначения, использованные нами для описания гармонического осциллятора, выбраны таким образом, что для описания полевых теорий требуется лишь небольшое обобщение. Гиперповерхность равного времени $\Sigma(a)$ становится некоторой общей пространственноподобной гиперповерхностью. Конфигурация поля на Σ обозначается $q : \Sigma$, если ее следует рассматривать как функционал поля-истории q. В полевых теориях с ограничениями, таких, как ОТО, существует неоднозначность в понятии «поле-история». Является ли гравитационное поле-история пространственно-временным метрическим тензором, который может либо удовлетворять, либо не удовлетворять любым ограничениям, или тензором, который должен снимать эти ограничения? Мы будем просто использовать то определение, которое для данной проблемы является более удобным. В § 6 для обсуждения гравитационного рассеяния используется понятие пространственно-временного метрического тензора, а в § 7 применяется «сверхпространственная» картина, в которой некоторые ограничения уже сняты и которая более удобна для рассмотрения квантованных замкнутых геометрий.

Рассмотренный подход развивает описание квантовой теории гильбертова пространства с помощью векторов-историй. Возможна и желательна и обратная процедура, дающая описание в гильбертовом пространстве квантования с помощью векторов-историй, или сумм по историям. Главная идея заключается в построении удовлетворительного разложения единичного оператора гильбертова пространства. Когда эти разложения вносятся в скалярное произведение для гильбертова пространства, они дают соотношения для сумм по историям, подобные уравнению (5.33). Эта процедура была использована Клаудером [53] для определения применимой к системам Ферми — Дирака формулировки, использующей интегральные функционалы. Развитый здесь подход с помощью векторов-историй требует знания правильного выражения для квантового действия и структуры группы однородности, в то время как подход с точки зрения разложения единичного оператора требует знания квантовой операторной алгебры. Оба подхода полностью дополняют друг друга. Каждый определяет то, что другой допускает.

Рамки описания с помощью векторов-историй достаточно широки и допускают много различных интерпретаций. Правила Фейнмана сумм по историям получаются из функциональной интегральной формы скалярного произведения (5.33). Уравнения (5.36) и (5.37) являются исходными для схемы квантования канонического оператора Швингера. Если постулировать уравнение (5.37) в качестве *ограничения*, которое рассматривается как определяющее физическое подпространство $\mathscr{V}$, то получается схема «ковариантного квантования» Новожилова и Тулуба [70], обсуждавшаяся Клаудером [54, 55]. Физическое подпространство, определяемое этим способом, является лишь одним из многих функциональных представлений $\mathscr{H}$ в $\mathscr{V}$. Получаемое представление зависит от функционала действия, который входит в «ограничения» (5.37), и от выбора канонических переменных. Конечно, тот факт, что одна схема квантования может быть получена из другой, не является чем-то исключительным. Однако процесс установления подобных связей обычно сложен и не дает новой информации. На языке векторов-историй преобразования весьма просты и полностью раскрывают отношение одной схемы к другой.

Подход, использующий векторы-истории, мы предлагаем главным образом как основу для установления связи между различными техниками квантования, применяемыми в ОТО. Однако наш подход не является всего лишь формальным новым изложением квантовой теории. Если принять, что *каждый* вектор-история обозначает квантовое состояние в $\mathscr{H}$, то можно легко строить «странные» квантовые теории путем выбора «странных» функциональных представлений. Такого рода нетривиальные утверждения на языке векторов-историй понадобятся нам ниже в этом параграфе для того, чтобы включить теорию источников Швингера в нашу унифицированную схему. Ниже мы рассмотрим одно из таких «странных» представлений квантовой ОТО.

Основой для данного примера, а также для обсуждения теории источников в конце этого параграфа служит определенный тип представления, который мы назовем представлением классических источников (КИ-представлением). Говорят, что функционал-история $F[q]$ принадлежит к КИ-представлению $\mathscr{s}(\Sigma)$, если он, во-первых, обращается в нуль, если q не удовлетворяет заданным уравнениям движения, и, во-вторых, зависит от q только через свою конфигурацию на гиперповерхности равного времени Σ. Заданные уравнения движения являются совершенно произвольными. Для удобства будем называть историю, которая им удовлетворяет, «классической историей». Странная природа КИ-представления становится очевидной при попытке интерпретировать состояния, описываемые вектором-историей. Для F в $\mathscr{s}(\Sigma)$ состояние $|F\rangle$ может быть отождествлено только с экспериментом, допускающим лишь классические истории. Такого эксперимента, конечно, не существует, и поэтому отсутствует прямая физическая интерпретация $|F\rangle$. Способ интерпретации КИ-векторов-историй заключается в том, чтобы записывать их в форме

$$F(q) = \Psi[q:\Sigma]\,\chi(q),$$

где $\chi(q) = 1$, когда q — классическое, и $\chi(q) = 0$ в других случаях. Функционал Ψ является функционалом состояния, который представляет состояние F. В конечномерных системах Ψ является функцией состояния.

Основное преимущество КИ-представления заключается в том, что пропагационный функционал $\langle\delta[(q-q_1):\Sigma']|\delta[(q-q_0).\Sigma]\rangle$, который связывает состояния с точно описанной конфигурацией на разных гиперповерхностях, может быть оценен без построения интегрального функционала. Выражение в виде суммы по историям (5.33) для этого скалярного произведения содержит только классическую историю, связывающую $q_0 : \Sigma$ с $q_1 : \Sigma'$. Непосредственно получаем

$$K[q_0 : \Sigma,\ q_1 : \Sigma'] = \text{const} \times \left\{ \exp \frac{i}{\hbar} I[\mathfrak{s}(\Sigma),\ \mathfrak{s}(\Sigma'),\ q_{\text{класс}}] \right\}. \quad (5.40)$$

Этот результат одновременно вскрывает основной недостаток КИ-представления. Функционал $I[\mathfrak{s}(\Sigma),\ \mathfrak{s}(\Sigma'),\ q]$ в общем случае *не является* классическим действием.

Рассмотрим уравнение (5.36), которое первоначально привело к отождествлению фазового функционала I с классическим действием. Поскольку классическая история, связывающая две конфигурации, существенно единственна, вариация $\delta_{\mathfrak{s}(\Sigma) \times \mathfrak{s}(\Sigma')}$ просто не существует. Следовательно, уравнение (5.36) в КИ-представлении является пустым. Поэтому нет серьезных оснований считать I функционалом действия. Критерием для выбора I в КИ-представлении служит то, что выражение (5.40) должно давать правильный пропагатор. В линейной полевой теории удается выбрать I так, чтобы он был классическим действием. При наличии взаимодействий I в лучшем случае является лишь приближением к классическому действию.

КИ-представление дает метод угадывания полного пропагационного функционала (S-матрицы, если взят подходящий предел), требующий лишь минимального внимания к промежуточным процессам. Этот метод лежит между теорией поля, которая рассматривает все промежуточные процессы, и S-матричным подходом, который допускает только классические (на массовой оболочке) промежуточные процессы. Теория источников Швингера может быть описана как КИ-представление теории частиц. Можно построить много различных КИ-представлений одной и той же физической системы. Для этого достаточно выбрать различные «классические уравнения движения» и подобрать фазовый функционал КИ-представления I так, чтобы получаемый при этом пропагационный функционал был правильным. Эта свобода соответствует в некоторой степени произвольности интерполяционного поля теории S-матрицы.

В качестве первого примера КИ-представления рассмотрим идею о том, что *все* фазовые интерференционные эффекты квантовой теории могут быть обусловлены гравитационным полем. Для *любой* системы полей и частиц фазовый функционал $I[\mathfrak{s}(\Sigma),\ \mathfrak{s}(\Sigma'), q]$ должен зависеть только от пространственно-временной геометрии, связывающей гиперповерхности Σ и Σ'. Требуется найти форму для I и систему уравнений движения, которые могут быть введены в уравнение (5.40), и дать правильный пропагационный функционал для *всей системы* частиц и полей.

Наше первое предположение об I основано на тех же представлениях о простоте, которые приводят к обычному функционалу действия ОТО. Пусть

$$I = \hbar \ell^2 \int_{\Sigma}^{\Sigma'} {}^4\sigma \{{}^4R\}, \tag{5.41}$$

где ℓ — постоянная длина, подлежащая определению. Наше первое предположение относительно уравнений движения состоит в том, что они являются обычными классическими уравнениями движения

$$G^{\mu\nu}(g_{\alpha\beta}) = \frac{8\pi k}{c^3} T^{\mu\nu}(g_{\alpha\beta}, \phi_i),$$

где все негравитационные переменные обозначены через ϕ_i, причем предполагается, что они подчиняются своим классическим уравнениям движения

$$E^i(\phi_i, g_{\alpha\beta}) = 0.$$

Полная система уравнений может быть получена из скалярного лагранжиана

$$\mathscr{L} = \frac{c^3}{8\pi k} {}^4R + \mathscr{L}_M(g_{\alpha\beta}, \phi_i), \tag{5.42}$$

где $\mathscr{L}_M$ — негравитационный лагранжиан обычной материи.

Теперь внесем это первое предположение о фазовом функционале и уравнениях движения в уравнение (5.40). Вследствие уравнений поля Эйнштейна полный лагранжиан на классических историях обращается в нуль, так что

$${}^4R = -\frac{8\pi k}{c^3} \mathscr{L}_M$$

и уравнение (5.40) дает

$$K[q:\Sigma, q:\Sigma'] = \text{const} \times \exp\left\{-i\left(\frac{8\pi k}{c^3\ell^2}\right)\int {}^4\sigma \mathscr{L}_M\right\}. \tag{5.43}$$

Если теперь выбрать

$$\ell = L = \left(\frac{8\hbar k}{c^3}\right)^{1/2} = 11{,}4 \cdot 10^{-33} \text{ см},$$

то уравнение (5.43) аппроксимирует пропагационный функционал, который можно получить обычным квантованием одних только негравитационных переменных.

Этот результат для пропагационного функционала имеет несколько недостатков, которые должны быть устранены путем второго приближения или исправления фазового функционала I и уравнений движения. Уравнение (5.43) всего лишь приближается к правильному негравитационному пропагационному функционалу, когда Σ' близко к Σ. В общем случае для получения правильного результата требуются чрезвычайно нелокальные исправления в уравнениях движения и фазовом функционале. Второй, более легко устранимый недостаток заключается в том, что этот результат не порождает гравитационных волн в асимптотических областях открытых вселенных. Выход из этого положения заключается в том, чтобы к первому приближению для фазового функционала прибавить поверхностный интеграл $\int dtE_\infty$, рассмотренный в § 4, п. 3.

Что же мы получили в результате таких рассуждений? Общепринято считать, что малость гравитационной квантовой длины $L \approx 10^{-33}$ см исключает сколько-нибудь значительные квантовые гравитационные эффекты. Но мы вывели квантовую длину, исходя из допущения, что *все* квантовые эффекты могут быть описаны как квантовые гравитационные по своей природе. Этот вывод не противоречит утверждению, что *собственно* квантовые гравитационные эффекты являются незначительными. Однако он показывает, что нельзя исходить только из малости квантовой длины и что получение такого определения собственно квантовых гравитационных эффектов, которое не смешивало бы их с непрямыми эффектами, может оказаться нелегкой задачей.

Наш пример КИ-представления может быть использован для иллюстрации одной из опасностей, связанных с использованием «странных» представлений, а именно того, что данное «представление» может оказаться вообще не представлением. Оно может не охватывать полное пространство физических состояний. Рассмотрим замкнутую *пустую* вселенную в описанном выше КИ-представлении. Нам известно, что $G^{\mu\nu} = 0$ и $\mathscr{L}_M = 0$, так что $I = 0$. Вспомнив функционал $\chi(q)$, равный единице для геометрий Эйнштейна и равный нулю в других случаях, можно из уравнения (5.35) заключить, что

$$\langle F|G\rangle = \langle F|\chi\rangle\langle\chi|G\rangle$$

для всех КИ-функционалов F и G. Все векторы-истории в КИ-представлении замкнутой пустой вселенной представляют одно и то же единственное физическое состояние, а именно $|\chi\rangle$. Учитывая его инвариантные свойства (независимость от выбора Σ), целесообразно назвать $|\chi\rangle$ вакуумным состоянием. В § 7 и 8 мы увидим, что идея о выделенном состоянии для квантованных замкнутых вселенных естественно возникает в любой общековариантной квантовой ОТО. Одна из возможных интерпретаций полученного результата состоит в том, что векторы-истории в КИ-представлении не обозначают все физические состояния и что существуют состояния, отличные от $|\chi\rangle$. Другая возможная интерпретация заключается в том, что действительно существует лишь одно состояние и наш результат правилен.

Могут ли «странные» представления приводить к стандартным предсказаниям квантовой теории поля? Ниже мы увидим, что по крайней мере одно из них, а именно теория источников Швингера, действительно дает правильные результаты.

4. Теория источников

Недавно Швингер применил еще один подход к квантованию гравитации, основанный на его «теории источников». Швингер вводит теорию источников как феноменологическую теорию частиц, имеющую мало общего с традиционной теорией поля. В основном мы будем рассматривать приложение теории источников к нашему объединенному подходу, а читателям, которые интересуются теорией источников, мы советуем обратиться к работам Швингера [84—86].

Фундаментальная концепция теории источников заключается в том, что состояние системы можно охарактеризовать заданием набора чисел заполнения и системы источников, рождающих и уничтожающих частицы. Одночастичные состояния с 4-импульсом p и источником $S(p)$ обозначаются как $|1_p\rangle^S$. Здесь источник рассматривается как амплитуда рождения

$$iS(p) = {}^S\langle 1_p | 0\rangle^S, \tag{5.44}$$

где элемент фазового объема понимается в виде $(d^4p/2p^0)(2\pi)^{-3}$. Пространственно-временная функция источника $S(x)$ получается из $S(p)$ преобразованием Фурье. Предполагается, что состояния можно рассматривать как сходящиеся или как расходящиеся (— или +) и что функция источника может быть соответственно разделена на начальную часть S_-, такую, что для всех сходящихся состояний

$$|\Psi_-\rangle^S = |\Psi_-\rangle^{S_-},$$

на конечную часть S_+, такую, что для всех расходящихся состояний

$$|\Psi_+\rangle^S = |\Psi_+\rangle^{S_+},$$

и часть взаимодействия S_i, такую, что

$$S = S_+ + S_- + S_i.$$

Предполагается, что начальная и конечная части являются предметами экспериментального контроля и наблюдений. Часть, ответственная за взаимодействие, является некоторым функционалом от S_- и S_+.

Теория источников вычисляет амплитуду вероятности для эксперимента по рассеянию, в котором источники S_- действуют на *вакуум*, рождая сходящиеся частицы, а источники S_+ поглощают все расходящиеся частицы, восстанавливая *вакуум*. Амплитуда такого эксперимента записывается в виде

$$\exp\{iW(S)\} = {}^S\langle 0_+ | 0_- \rangle^S.$$

Теперь следует определить вид W и найти зависимость S_i от S_+ и S_-. Процедура определения W, предложенная Швингером, основана на разложении тождественного оператора в представлении источников. Примем подход, основанный на векторах-историях, который, как мы упоминали выше, дополняет подход, основанный на разложении единичного оператора, путем перестановки местами некоторых выводов и исходных положений.

Откажемся от определения источника как амплитуды рождения и примем полевую картину, в которой состояния частиц описываются полем ϕ с квадратичным действием $I[\phi]$. Здесь ϕ имеет невыписанные индексы [тензорные, спинорные, $SU(3)$ и т. д.], соответствующие трансформационным свойствам и классификации рассматриваемой системы частиц. Предлагается рассмотреть КИ-представление квантовой теории, которое имеет $I[\phi]$ в качестве фазового функционала и уравнения движения в виде

$$\frac{\delta I}{\delta \phi} = S \tag{5.45}$$

с соответствующими связями на $S(x)$. Для описания на этом языке конкретной физической системы нужно выбрать подходящие трансформационные свойства для S и уточнить ограничения на S. В этой картине наиболее отчетливые физические интерпретации $S(x)$ обеспечиваются легко выводимыми результатами

$$\left\langle \frac{\delta I}{\delta \phi(x)} \right\rangle = {}^S\langle 0_- | T(\Sigma) \left[\frac{\delta I}{\delta \phi(x)} \right] | 0_- \rangle^S = S_-(x) \quad \text{для } x \text{ на } \Sigma,$$

$$\left\langle \frac{\delta I}{\delta \phi(x)} \right\rangle = {}^S\langle 0_+ | T(\Sigma') \left[\frac{\delta I}{\delta \phi(x)} \right] | 0_+ \rangle^S = S_+(x) \quad \text{для } x \text{ на } \Sigma'$$

для среднего значения классического источника $\delta I/\delta\phi$. Заметим, что это рассмотрение неприменимо в «зоне взаимодействия» между Σ и Σ', так как КИ-функционалы-истории не могут быть там интерпретированы.

Пусть теперь $V_{-}[\phi:\Sigma]$ и $V_{+}[\phi:\Sigma']$ представляют собой функционалы прошлого и будущего состояний вакуума *в конфигурационном представлении.* В этом случае КИ-представлением для вакуума будет $|V_{-}\chi\rangle$ в прошлом и $|V_{+}\chi\rangle$ в будущем. Здесь χ является функционалом, определенным в п. 3. Он равен единице, когда удовлетворяются уравнения движения, и равен нулю в других случаях. Для определения $W(S)$ необходимо из уравнения (5.33) или (5.40) найти скалярное произведение между прошлым и будущим вакуумными состояниями. Полученная при этом сумма по историям является тривиальной, так как, кроме обычных КИ-ограничений на классические истории, здесь ядро $\exp\{(i/\hbar)I[q_{\text{класс}}]\}$ может быть факторизовано. Выразим общее решение уравнения (5.45) в виде суммы решения ϕ_B однородных уравнений ($S=0$) и решения

$$\phi_S=\int\limits_{\Sigma}^{\Sigma'} d^4x' G(x, x') S(x')$$

(подразумевается свертка по индексам) уравнения (5.45), которое обращается в нуль на границах интегрирования $\int\limits_{\Sigma}^{\Sigma'} d^4x$. Квадратичное действие $I[q_{\text{класс}}]$ в этом случае становится простой суммой $I[q_B]+I[q_S]$, а уравнение (5.35) дает

$${}^{S}\langle 0_{+}|0_{-}\rangle^{S}={}^{0}\langle 0_{+}|0_{-}\rangle^{0}\exp\left\{\left(\frac{i}{\hbar}\right)I[q_S]\right\}.$$

Если вакуум устойчив при отсутствии источников, то скалярное произведение в правой части этого выражения равно единице, и мы заключаем, что

$$W(S)=\left(\frac{i}{\hbar}\right)I[q_S]=\left(\frac{i}{\hbar}\right)\int d^4x\int d^4x'\{S(x)\,G(x, x')\,S(x')\}. \quad (5.46)$$

Это фундаментальный формальный результат теории источников.

Чтобы от этого анализа на основе векторов-историй вернуться к полной теории источников Швингера, необходимо вывести уравнение (5.44), связывающее $S(p)$ с амплитудой рождения. Вывод получается прямолинейным, но неизящным, поэтому мы лишь наметим ход рассуждений. Сначала получают обычный оператор рождения в форме вариационной производной. Затем необходимо показать, что получаемый оператор, действующий на часть КИ-вектора-истории, содержащую функционал состояния, действительно порождает состояние частицы со всеми обычными трансформационными свойствами (это наиболее запутанная часть вывода). Затем обращают внимание на то, что правая часть уравнения (5.44), выраженная через оператор рождения, действующий на вакуум, может быть отождествлена с правой частью уравнения (5.37). Сравнив левые части этих двух уравнений, получают уравнение (5.44).

Чисто формальные результаты, рассмотренные выше, являются всего лишь небольшой частью теории источников. Для получения полного описания физической системы необходимо указать точные граничные условия, налагаемые на $G(x, x')$, а также трансформационные свойства S и G и способ сворачивания индексов в уравнении (5.46). Далее необходимо решить, какие ограничения должны быть наложены на функцию источника $S(x)$. При выборе указанных выше величин руководствуются требованиями явной лоренц-ковариантности, условием вероятности $|{}^{s}\langle 0_+|0_-\rangle^{s}| < 1$ и условием тождественности $W(S)$ с проинтегрированной по времени энергией при квазистатическом распределении источников. Лоренц-ковариантная форма G имеет мнимую часть, которая ставит под сомнение вероятностное условие, если только мнимая часть $W(S)$ сама не является положительно определенной. Обычно это требование приводит к закону сохранения, которому подчиняется $S(x)$.

Для введения взаимодействия выражают функциональную зависимость S_i от S_+ и S_-. Это производится путем принятия допущения о «единственности источника», согласно которому S_i является простой аналитической функцией полного источника S. Таким образом,

$$S_i = V[S] = V(S_+ + S_i + S_-). \tag{5.47}$$

Затем для отыскания решения уравнения связи (5.47) строится итерационная последовательность:

$$S_i^{(0)} = V(S_+ + S_-),$$

$$S_i^{(1)} = V(S_+ + S_i^{(0)} + S_-) = V\{S_+ + V(S_+ + S_-) + S_-\},$$

$$S_i^{(n+1)} = V(S_+ + S_i^{(n)} + S_-).$$

Первый член этой последовательности называют «первичным взаимодействием».

Для демонстрации способа описания взаимодействий в теории источников рассмотрим

$$S_i(\xi) = V(\xi)[S] = \int d^4x \int d^4x' \int d^4x'' G_3(x, x', x'', \xi) S(x) S(x') S(x'').$$

Чтобы выявить все члены, входящие в первое приближение для S_i, опустим все интегралы. Тогда

$$S_i^{(0)} = G_3[S_+S_+S_+ + S_+S_+S_- + S_+S_-S_+ + S_+S_-S_- + \\ + S_-S_+S_+ + S_-S_+S_- + S_-S_-S_+ + S_-S_-S_-].$$

Каждый из этих членов называют «эффективным источником». Когда такой источник $S_{\text{эфф}}$ входит в $W(S)$, он порождает члены вида $S_\pm(x) G(x, x') S_{\text{эфф}}(x')$, которые интерпретируются как поглощение приходящей частицы или рождение уходящей частицы эффективным источником. Так, $S_-GG_3S_+S_+S_+$ интерпретируется как поглощение приходящей частицы эффективным источником, мощность которого зависит от амплитуды, определяющей исходящее трехчастичное состояние. Этот член говорит о том, что приходящая частица поглощается только тогда, когда наблюдаются три уходящих частицы (продукты распада). Другие члены могут быть интерпретированы подобным же образом. При использовании этой интерпретации становится ясным, что свойство перекрестной симметрии появляется чисто алгебраически и является следствием исходного допущения о единственности источника.

Как найти первичное взаимодействие? Это делается путем наложения очевидных формальных требований простоты и лоренц-ковариантности, некоторых менее очевидных формальных требований, которые мы не будем рассматривать, и, наконец, как указал Швингер, путем эксперимента.

В случае гравитационного поля подходящая форма $W(S)$ имеет вид

$$W = \frac{1}{2} \int d^4x \int d^4x' \left[S^{\mu\nu}(x) D_+(x - x') S_{\mu\nu}(x') - \frac{1}{2} T(x) D_+(x - x') T(x') \right],$$

где

$$S \equiv \gamma_{\mu\nu} S^{\mu\nu} \quad (\gamma_{\mu\nu} \text{ — метрика Минковского}),$$

$$S_{\mu\nu} \equiv \gamma_{\mu\alpha}\gamma_{\nu\beta} S^{\alpha\beta},$$

а $S^{\mu\nu}$ подчиняется закону сохранения

$$S^{\mu\nu}_{,\nu} = 0.$$

Эта форма $W(S)$ порождает ньютоновский потенциал для квазистатических источников и, будучи рассмотренной совместно с положениями специальной теории относительности, дает правильные предсказания красного смещения и упреждения перигелия [74, 86]. Гравитационное самодействие может быть получено с помощью требования

$$S_i^{\mu\nu} = \frac{\delta W(S)}{\delta \gamma_{\mu\nu}},$$

которое означает, что энергия-импульс поля является также его источником.

Заметим, что $S^{\mu\nu}$ не является просто общековариантным тензором энергии-импульса $T^{\mu\nu}$. Она включает также некую форму гравитационного псевдотензора энергии-импульса, которую часто обозначают $t^{\mu\nu}$. Использование собственной энергии-импульса поля в качестве источника может быть обосновано несколькими способами. Самая простая точка зрения основана на том, что наличие связи между гравитацией и энергией-импульсом материи считается уже известным. Если бы гравитация не была связана с полной энергией-импульсом, в которую входит ее собственная энергия, то полный тензор источника $S^{\mu\nu}$ не сохранялся бы в случаях, когда гравитационная энергия-импульс сообщается веществу. Тогда нарушалась бы вероятностная интерпретация вакуум-вакуумной амплитуды. Другая точка зрения исходит из того, что поле $h_{\mu\nu} \equiv \delta W / \delta S^{\mu\nu}$, содержащее гравитационные состояния, представляет собой калибровочную группу, которая может быть отождествлена с инфинитезимальными преобразованиями координат. Естественное требование того, чтобы полная теория оставалась инвариантной при произвольном преобразовании координат, приводит к взаимодействию в теории Эйнштейна [86].

Тот факт, что $S^{\mu\nu}$ не является энергией-импульсом материи, является фундаментальной трудностью такого подхода к ОТО. В общем случае находят, что

$$S^{\mu\nu} = s^{\mu\nu} + T^{\mu\nu},$$

где $s^{\mu\nu}$ — чисто гравитационный член, а $T^{\mu\nu}$ — тензор материи, подчиняющийся ковариантным законам сохранения

$$T^{\mu\nu}_{;\nu} = 0.$$

Для вычисления амплитуды рассеяния в конкретном эксперименте нужно найти полностью сохраняющуюся S_-, которая соответствует гравитационному генератору, и полностью сохраняющуюся S_+, представляющую приемник. Эта проблема далеко не тривиальна, Нужно извлечь контролируемые параметры материи из ковариантно сохраняющейся части $S^{\mu\nu}$. Швингер предложил решение этой проблемы на основе построения модельных источников для гравитонов с помощью теории возмущений.

Используя теорию источников гравитонов, Радковский и Швингер вычислили поправку наинизшего порядка к эффективному пропагатору, а также соответствующую поправку к ньютоновскому потенциалу. Главным результатом является член $1/r^3$ в потенциале, который может быть приписан пространственной неопределенности в r порядка 10^{-33} см.

В наш беглый обзор теории источников не вошло несколько важных аспектов этой теории. Мы не рассмотрели трактовку в теории источников спина и статистики частиц и античастиц. Мы также опустили обсуждение расходимостей. Поскольку, по мнению некоторых авторов, избежание расходимостей представляет собой главную идею теории источников, мы сделаем два замечания, которые по существу объясняют избежание расходимостей. Прежде всего итерационная процедура для определения S_i не всегда сохраняет разделение S_+ и S_- от S_i. В некоторых случаях необходимо применять спектральную нормализацию на каждом шаге итерации для обеспечения того, чтобы полный источник поглощал и выделял одни и те же начальные и конечные состояния. Наше второе замечание состоит в том, что «пузырьковый» член вида $S_{\text{эфф}}GS_{\text{эфф}}$ описывает рождение группы частиц в одной точке и ее поглощение в другой. Отметим, что для всей группы частиц имеется только *один* пропагатор. Вся группа виртуальных частиц рассматривается как одна частица с 4-импульсом и квантовыми числами, которые являются суммами 4-импульсов и квантовых чисел отдельных частиц. Процессы, в которых несколько частиц независимым образом переходят от общей точки рождения к общей точке гибели, просто не могут иметь места в теории источников. Поскольку именно такие процессы вызывают расходимости, теория источников обходит проблему расходимости.

§ 6. ГРАВИТОННОЕ РАССЕЯНИЕ: СОВРЕМЕННЫЕ РЕЗУЛЬТАТЫ И ПРОГНОЗЫ

Построение амплитуд гравитонного рассеяния является особой, но крайне важной областью квантовой ОТО. Особой она является потому, что сосредоточивает свое внимание на весьма ограниченном множестве процессов в открытых вселенных. Важность же ее для квантовой ОТО заключается в том, что она привычным образом рассматривает трудности полной теории, связанные с расходимостями. Таким образом, имеется хорошо определенная физическая проблема (вычисление S-матрицы), а также приложение к «основному направлению» физики — устранению серьезных расходимостей в квантовой теории поля. По этим причинам гравитонное рассеяние стало наиболее активной областью современных исследований в квантовой ОТО.

Чтобы представить в рамках единого подхода основные понятия и уравнения теории гравитонного рассеяния, мы используем формализм векторов-историй. В наши намерения не входит полное рассмотрение технических деталей. Скорее мы попытаемся дать общую схему, которая коснется важнейших положений и обеспечит читателю достаточное количество технических деталей для овладения, если он того пожелает, соответствующим математическим аппаратом.

Если исходить из логических соображений, то редукционный формализм, используемый для выделения физической информации, должен предшествовать чисто формальному обсуждению обобщенных функций Грина. Однако мы будем следовать в противоположном направлении, так как чисто формальное обсуждение можно производить, не ограничиваясь открытыми асимптотически симметричными пространствами-временами, чего, по-видимому, требует редукционный формализм. Подразумевает ли такое ограничение в нашем современном представлении о редукционном формализме фундаментальную ограниченность S-матричной точки зрения на открытые вселенные? Этот вопрос не решен. Мы придерживаемся осторожного подхода, не исключающего S-матрицы замкнутой вселенной до тех пор, пока это не становится абсолютно необходимым.

1. Обобщенные функции Грина

В настоящем обсуждении принимается, что гравитационная история q представляет собой пространственно-временную геометрию, регулярную всюду, кроме, возможно, изолированных поверхностей разрыва и точечных сингулярностей. Предполагается, что все пространства-времена имеют одинаковую топологию. Эта топология должна быть совместной с существованием полных однопараметрических семейств пространственноподобных гиперповерхностей — каждая пространственно-временная точка пересекается по крайней мере одной гиперповерхностью $\Sigma(t)$ из каждого семейства σ. Наконец, каждая пространственно-временная геометрия представляется компонентами $g_{\mu\nu}$ метрического тензора, рассматриваемыми как функции на *опорном пространстве-времени* q_0. Опорными метрическими компонентами являются $\gamma_{\mu\nu}$. Опорная или фоновая геометрия по существу является формальным приемом, применяемым для рассмотрения калибровочных (координатных) проблем. Этот прием использовал де Витт [26]. Ковариантные производные по отношению к q_0 мы определяем как

$$A_{\mu\cdot\nu} \equiv A_{\mu,\,\nu} - A_\alpha \Gamma^\alpha_{\mu\nu}(q_0),$$

где

$$\Gamma^\alpha_{\mu\nu}(q_0) \equiv \frac{1}{2}\gamma^{\alpha\beta}(\gamma_{\beta\mu,\,\nu} + \gamma_{\beta\nu,\,\mu} - \gamma_{\mu\nu,\,\beta}).$$

Для S-матричных расчетов предполагаем, что физическое гильбертово пространство $\mathscr{H}$ натянуто на векторы $|F_{\sigma\pm}\rangle$, где σ — семейство пространственноподобных гиперповерхностей, а $F_{\sigma\pm}$ — функционал-история

$$F_{\sigma\pm}(q) = \lim_{t\to\pm\infty} F[g_{\mu\nu} : \Sigma(t)]. \tag{6.1}$$

Гиперповерхности $\Sigma(t)$ определяются геометрически через историю q. Они могут быть поверхностями $x^0 = \mathrm{const}$, если координаты x^0 фиксированы дополнительным условием. В общем случае эти векторы преобразуются сложным образом при изменении семейства σ (конечно, если это изменение не действует на $\Sigma(t)$, только когда t принадлежит некоторому конечному интервалу). Векторы $|F_{\sigma-}\rangle$ являются начальными состояниями, а векторы $|F_{\sigma+}\rangle$ — конечными. Мы будем предполагать, что существует по крайней мере одно начальное состояние, инвариантное относительно всех преобразований в семействе σ. Это состояние мы запишем в виде

$$|-\rangle = |V^{(-)}_{\sigma-}\rangle \tag{6.2}$$

и назовем его начальным вакуумом. Аналогично мы предполагаем существование конечного вакуума, обозначаемого $V^{(+)}_{\sigma+}$. Связь между $|-\rangle$ и $|+\rangle$ и между функционалами $V^{(-)}(q)$ и $V^{(+)}(q)$ остается открытой.

Скалярное произведение в $\mathscr{H}$ записывается в форме интегрального функционала (5.33):

$$\langle F_{\sigma+}|G_{\sigma-}\rangle = \int_{\sigma-}^{\sigma+} \mathscr{D}q\,[F^*_{\sigma+}(q)\,G_{\sigma-}(q)\exp\{iI(q)\}]. \tag{6.3}$$

Действие $I(q)$ необходимо выбрать так, чтобы вариации δ_{-+}, в которых $g_{\mu\nu} : \Sigma(\pm\infty)$ фиксировано, давали уравнения поля без остаточных поверхностных членов. Действие, удовлетворяющее этому требованию, есть $I_{\text{ГГ}}$ (см. табл. 1). Другой подходящей формой действия является

$$I=\int_{-\infty}^{\infty}dt\left\{\int\sigma\left(\pi^{ij}\dot{g}_{ij}-N_{\mu}R^{\mu}\right)-E_{\infty}\right\},\tag{6.4}$$

где E_{∞} — поверхностный интеграл (4.11), который необходимо включить, чтобы гамильтониан был равен полной энергии гравитационного поля. Здесь мы находим другую мотивировку для включения поверхностного интеграла — он исключает остаточные поверхностные члены, которые возникают при варьировании скалярной 3-кривизны ${}^3R$, входящей в R^0.

Операция $\int\mathscr{D}q$ интерпретируется как суммирование по пространственно-временным геометриям (как в ковариантном квантовании, рассмотренном Клаудером [54, 55]). В идеале эту операцию следовало бы производить, обеспечив сначала единственное и несингулярное представление для каждой геометрии, а лишь затем суммируя представления. К сожалению, приемлемая схема представления не известна. В качестве компромисса можно предложить следующую процедуру (по существу та же процедура использовалась де Виттом [25]).

1. Представляем каждое пространство-время метрическим тензором, удовлетворяющим четырем дополнительным условиям $C^{\mu}(g_{\alpha\beta})=0$, таким, как

$$C^{\mu}=\left(g^{\mu\alpha\cdot\beta}-\frac{1}{2}\gamma^{\mu\alpha}g^{\rho\sigma\cdot\beta}\gamma_{\rho\sigma}\right)\gamma_{\alpha\beta}=0.\tag{6.5}$$

Эти условия выбраны так, чтобы обеспечить несингулярность представлений для пространств-времен вблизи q_0 (т. е. в окрестности первого порядка q_0).

2. Суммируем по $g_{\mu\nu}$, удовлетворяющим дополнительным условиям для данной опорной геометрии q_0.

3. Для обеспечения корректной трактовки сингулярных представлений, генерируемых дополнительными условиями, требуем, чтобы все физические результаты не зависели от опорного метрического тензора $\gamma_{\mu\nu}$.

Компоненты опорной метрики можно изменять, либо меняя геометрию q_0, либо изменяя координаты в q_0. Инвариантность относительно изменений опорных координат можно обеспечить, записывая все в явно ковариантных относительно q_0 обозначениях. Инвариантность относительно изменений геометрии q_0 должна исследоваться в явном виде.

В дополнение к представлению каждой пространственно-временной геометрии требуем существования группы однородности для $\int \mathscr{D}q$. Для линейных дополнительных условий, которые мы приводим в качестве примера, следует выбрать преобразования

$$g^{\mu\nu} \to g^{\mu\nu} + a^{\mu\nu},$$

где $a^{\mu\nu}$ не зависит от q. Однако можно выбрать нелинейные дополнительные условия вместе с неаддитивной группой однородности, в которой $\delta g^{\mu\nu}$ зависит от q. С этим вопросом связана также возможность наличия некоторой вещественной плотности, на которую умножается однородный интегральный функционал, как в выражении

$$\int \mathscr{D}q \ldots = \int \mathscr{D}g^{\mu\nu}\Delta(q) \ldots .$$

Эту плотность можно учесть, добавляя $-\ln \Delta$ к действию $I(q)$.

Мы считаем удобным принять сокращенные обозначения де Витта, в которых латинский индекс обозначает пару пространственно-временных индексов вместе с четырьмя пространственно-временными координатами [25, 27]. Латинские индексы, появляющиеся вверху и внизу в одном и том же произведении, обозначают суммирование одновременно по дискретным и по непрерывным компонентам:

$$A_i B^i \equiv \int {}^4\sigma A_{\mu\nu} B^{\mu\nu}.$$

В этих обозначениях мы определяем обобщенные функции Грина как

$$G^{i_1 i_2 \ldots i_n} \equiv \frac{\langle + \mid T(\sigma -)\left[h^{i_1}_{\cdot} h^{i_2}_{\cdot} \ldots h^{i_n} \right] \mid - \rangle}{\langle + \mid - \rangle}, \tag{6.6}$$

где $T(\sigma-)$ $[\ldots]$ — хронологическое произведение, определяемое уравнением (5.35), а

$$h^i = \dot{h}^{\mu\nu}(P) = g^{\mu\nu}(P) - \gamma^{\mu\nu}(P).$$

Мы будем использовать также «сверхсокращенные» обозначения де Витта, которые заменяют набор индексов $i_1, i_2, \ldots, i_n$:

$$G^{i_1 i_2 \ldots i_n} = G^{(n)}.$$

Когда сверхсокращенные индексы повторяются сверху и снизу в одном и том же произведении, это означает суммирование по целым значениям n от 1 до ∞, а также по всем скрытым индексам $i_1, i_2, \ldots, i_n$ для каждого значения n. Таким образом,

$$A_{(n)}B^{(n)} = A_i B^i + A_{ij}B^{ij} + A_{ijk}B^{ijk} + \ldots = \int {}^4\sigma \left[A_{\mu\nu}B^{\mu\nu}\right] +$$
$$+ \int {}^4\sigma \int {}^4\sigma' \left\{A_{\mu\nu\mu'\nu'}(P, P')\, B^{\mu\nu\mu'\nu'}(P, P')\right\} +$$
$$+ \int {}^4\sigma \int {}^4\sigma' \int {}^4\sigma'' \left\{A_{\mu\nu\mu'\nu'\mu''\nu''}(P, P', P'')\, B^{\mu\nu\mu'\nu'\mu''\nu''}(P, P', P'')\right\} + \ldots .$$

Дальнейшее заимствование у де Витта — обозначение функциональной производной

$$A_{,\,i} = \frac{\delta A}{\delta g^i} = \frac{\delta A}{\delta g^{\mu\nu}(P)},$$
$$A_{,\,(n)} = A_{,\,i_1,\,i_2,\,\ldots,\,i_n}.$$

Для произведений типа $\delta g^{\mu\nu}(P)\,\delta g^{\mu'\nu'}(P')\ldots$ и $h^i h^j h^k$ мы пишем $(\delta g)^{(n)}$ и $(h)^{(n)}$ или просто $h^{(n)}$.

Основная цель, которую преследуют сверхсокращенные обозначения, — обеспечить компактную запись выражений

$$A(q + \delta q) = \frac{1}{r!} A_{,\,(r)}(\delta q)^{(r)} \tag{6.7}$$

и

$$A(q) = \frac{1}{n!} A_{,\,(n)}(q_0)\, h^{(n)} \tag{6.8}$$

для функциональных разложений Тейлора.

Для определения обобщенных функций Грина проинтегрируем по частям тождества (5.34), воспользуемся уравнением (6.3) и получим условие

$$\langle + \,|\, \frac{1}{n!} T(\sigma -)\left[\exp(-iI)\{\exp(iI)\}_{,\,(n)}(\delta_{+-}g)^{(n)}\right] |\, - \rangle = 0.$$

Затем, разлагая функционал в скобках в функциональный ряд Тейлора в окрестности q_0 и деля на $\langle + | - \rangle$, получаем

$$\frac{1}{n!} \exp(-iI)\{\exp(iI)\}_{,\,(n)}(\delta_{+-}g)^{(n)} +$$
$$+ \frac{1}{n!\,r!}\left[\exp(-iI)\{\exp(iI)\}_{,\,(n)}(\delta_{+-}g)^{(n)}\right]_{,\,(r)} G^{(r)} = 0, \tag{6.9}$$

где все функционалы вычисляются в q_0. Поскольку вариация $\delta_{+-}g^i$ произвольна (но не обязательно независима от q), это по существу совокупность уравнений, из которых можно вычислить обобщенные функции Грина. Для локального функционала действия, такого, как (6.4), эта совокупность уравнений представляет собой систему линейных дифференциальных уравнений в частных производных, в которых функции Грина более высокого порядка служат источниками для функций Грина низшего порядка. *Уравнение* (6.9) — *по существу математическая формулировка теории гравитонного рассеяния.* Мы получили этот результат прямо из подхода, использующего интегральный функционал. Чаще же эту совокупность уравнений выводят более простым способом по аналогии с обычной квантовой теорией поля. Интегральные функционалы тогда вводятся просто как удобный способ получения этой совокупности уравнений (см. [15], стр. 227—235, 421—430, 484—503; [25, 51]).

Для решения уравнений (6.9) следует прибегнуть к методу итерации. В этом методе первый шаг состоит в отбрасывании всех «членов взаимодействия», в которых функции Грина более высокого порядка служат источниками для функций Грина низшего порядка. В результате получается усеченная система, которая, конечно, эквивалентна системе, получаемой из квадратичного функционала действия линеаризованной теории. Функции Грина нулевого порядка $G_0^{r)}$, являющиеся решениями этой системы, можно получить, если представить $G_0^{(r)}$ в виде сумм по всем различным произведениям $G_0^{(1)}$ и $G_0^{(2)}$, которые вычисляются в r точках. Усеченная система тогда сводится к уравнению

$$I_{,i} G_0^i = 0,$$

показывающему, что G_0^i есть решение линеаризованных уравнений, или уравнений свободного поля, и уравнению

$$I_{,i,j} I_{,s,p} G_0^{js} = i I_{,i,p},$$

показывающему, что G_0^{js} является функцией Грина (с точностью до продольных частей) для свободного поля.

Выражение для обобщенных функций Грина в нулевом порядке имеет точно такую же структуру, как разложение Вика для хронологического произведения [15, 51, 82]. Там, где мы писали $T(\sigma—)$ [...], можно читать T[...], не делая какой-либо ошибки, хотя последнее выражение трудно с уверенностью определить в ОТО.

Следующий шаг в решении уравнений (6.9) состоит в подстановке $G_0^{(r)}$ в члены взаимодействия, $G_1^{(r)}$ — в линейные члены или члены свободного поля и решении полученной системы линейных дифференциальных уравнений в частных производных относительно $G_1^{(r)}$, рассматривая G_0^{il} как функции Грина. На этом этапе появляется трудность, заключающаяся в том, что все члены взаимодействия содержат обобщенные функции Грина с совпадающими точками вычисления. Система, определяющая $G^{(r)}$, не фиксирует отдельные значения, такие, как $G^{\mu\nu\alpha\beta}(P, P)$. Таким образом, относительно этих величин следует сделать отдельное предположение. Опуская пространственно-временные индексы, получаем

$$G(P_1, P_2, \ldots, P, P) = \lim_{P' \to P} G(P_1, P_2, \ldots, P, P'), \qquad (6.10)$$

что приводит к обычному выражению теории возмущений для обобщенных функций Грина. С помощью этого предположения можно, по крайней мере формально, найти $G_1^{(r)}$, подставить этот результат в члены взаимодействия, найти $G_2^{(r)}$ и т. д. Эта процедура наилучшим образом описывается с помощью фейнмановских правил вычисления радиационных поправок для пропагаторов ($G^{(2)}$) и вершинных функций [т. е. $\delta^N I/(\delta g)^{(N)}$]. Как обычно, в таких вычислениях возникают расходимости, по существу обусловленные уравнением (6.10). Тот факт, что ни метод итераций, ни уравнение (6.10) не являются основой для фундаментальной системы уравнений (6.9), оставляет место для точных решений (например, для вычисления в явной форме интегрального функционала), свободных от расходимостей.

2. *Редукционный формализм*

Чтобы извлечь информацию из обобщенных функций Грина, построим функционалы $F(\sigma+, r_+, q)$ и $F(\sigma-, r_-, q)$, такие, что векторы

$$|r_\pm\rangle = |F(\sigma\pm, r_\pm)\rangle$$

ортогональны в том смысле, что для некоторой меры μ и любой пробной функции f

$$\int d\mu(r_-)\left[\frac{\langle s_- | r_-\rangle}{\langle - | - \rangle}\right] f(r_-) = f(s_-)$$

(и аналогично для r_+ и s_+), полны в том смысле, что

$$\int d\mu(r_-) \int d\mu(r_+)\, |r_+\rangle \left[\frac{\langle r_+ | r_-\rangle}{(\langle + | + \rangle \langle - | - \rangle)}\right] \langle r_- |$$

является разложением тождественного или единичного оператора в квантовом гильбертовом пространстве, и *экспериментально наблюдаемы* в том смысле, что начальные состояния $|r_-\rangle$ соответствуют определенным рецептам подготовки экспериментов, а конечные состояния $|r_+\rangle$ соответствуют определенным рецептам интерпретации результатов измерений.

Если заданы состояния $|r_\pm\rangle$, то вероятность $P[B(+), A(-)]$ получения результата в наборе $B(+)$ из измерения состояния в наборе $A(-)$ есть

$$P[B(+), A(-)] = \int\limits_{B(+)} d\mu(r_+) \int\limits_{A(-)} d\mu(r_-) \mathscr{P}(r_+, r_-),$$

где плотность $\mathscr{P}$ равна

$$\mathscr{P}(r_+, r_-) = \left| \frac{\langle r_+ | r_- \rangle}{\langle - | - \rangle} \right|^2.$$

Комплексные числа

$$S(r_+, r_-) = \frac{\langle r_+ | r_- \rangle}{\langle - | - \rangle}$$

являются элементами матрицы рассеяния. Чтобы выразить эту матрицу через обобщенные функции Грина, определенные уравнением (6.6), разложим $[V_\pm^{(\pm)}]^{-1} F(r_\pm)$ в функциональный ряд Тейлора в точке q_0; тогда

$$S(r_+, r_-) = \frac{1}{n!} \{V_+^{(+)-1} F(\sigma+, r_+)\}_{,(n)} \times$$
$$\times \{V_-^{(-)-1} F(\sigma-, r_-)\}_{,(m)} G^{(n)(m)} \frac{\langle + | - \rangle}{\langle - | - \rangle}.$$

Это соотношение является формулировкой редукционного формализма.

Решающим шагом в редукции является построение внешних векторов состояния $|r_\pm\rangle$. Единственный известный эффективный путь построения этих векторов — ограничиться открытыми пространственноподобными гиперповерхностями $\Sigma(t)$ и предположить, что все пространственно-временные истории обладают асимптотической симметричной группой. Тогда допустимые семейства гиперповерхностей ограничиваются только гиперповерхностями, которые асимптотически приближаются к поверхностям постоянного времени, естественно обладающим асимптотической группой. Внутренние конечновременные вариации семейства σ несущественны для внешних состояний, поэтому допустимое множество семейств гиперповерхностей сокращается до множества с конечным числом параметров (эти параметры являются как раз параметрами асимптотической группы). Искомыми состояниями $|r_\pm\rangle$ будут состояния, преобразующиеся по неприводимым представлениям асимптотической группы. Эти состояния, конечно, являются *состояниями частиц*. Они представляют наборы *гравитонов*.

Один способ построения гравитонных состояний для частной асимптотической симметрии заключается в приведении матрицы

$$U(r_-, r'_-) = \langle r'_-, \sigma' | r_-, \sigma \rangle = \\ = \int_{\sigma-}^{\sigma'-} \mathscr{D}q \left[F_-(\sigma-, r_-, q) F^*_-(\sigma'-, r'_-, q) \exp\{iI(q)\} \right]$$

и аналогично матрицы $U(r_+, r'_+)$, насколько это возможно. Приведенная матрица содержит везде нулевые элементы, кроме набора подматриц вдоль диагонали. Интегральный функционал, определяющий U, полностью вычисляется в асимптотической области, поэтому можно использовать квадратичное действие линеаризованной теории. Нормальные моды этого действия представляют собой как раз тензорные поля, преобразующиеся по неприводимым представлениям асимптотической группы. Действие, выраженное через амплитуды этих полей, является суммой независимых действий гармонических осцилляторов. Тогда приведение U сводится к выражению $F(\sigma-, r_-, q)$ в виде подходящей комбинации собственных функций гармонических осцилляторов. В результате получаем функционалы $F(\sigma\pm, r_\pm)$, которые являются функционалами внешних состояний, соответствующих $|r_\pm\rangle$.

После того как получены матричные элементы между состояниями $|r_-\rangle$ и $|r_+\rangle$, можно использовать обычные методы получения сечений процессов гравитонного рассеяния. Такие сечения вычислили Вайнберг [89], а также де Витт [26].

3. Ковариантность, унитарность и фиктивные частицы

Когда асимптотическая симметрия является группой Пуанкаре (или по крайней мере включает ее), метод построения амплитуд гравитонного рассеяния представляет собой большей частью непосредственное обобщение диаграммных правил Фейнмана. Такое обобщение впервые было сделано самим Фейнманом [38]. Фейнман не пытался явно показать ковариантность относительно преобразований координат в опорной геометрии (в его рассмотрении это плоское пространство). Он просто отметил, что можно собрать диаграммы в частные суммы — *скобки Фейнмана*, которые калибровочно (координатно) инварианты.

Чтобы получить скобки Фейнмана, предположим, что набор входящих и выходящих волн классического гравитационного поля определяется координатно независимым способом. Такое определение возможно в асимптотической области [22]. Пусть амплитуды классических волн пропорциональны параметрам $\eta_1, \eta_2, \dots, \eta_N$; соответствующую классическую историю гравитационного поля обозначим $q_c(\eta_1, \dots, \eta_N)$. Функционал действия $B(\eta) = I\{q_c(\eta)\}$, вычисленный по этой истории, есть по построению координатно независимый объект. Если функцию $B(\eta)$ разложить в ряд по $\eta_1, \dots, \eta_N$, то каждый *коэффициент* разложения в отдельности будет координатно независим. Эти коэффициенты есть *примитивные скобки Фейнмана.*

Примитивные скобки Фейнмана можно представить в виде конечных сумм диаграмм Фейнмана, разлагая их по классической теории возмущений. Умножим неквадратичную часть I на ε, чтобы получить $I[\varepsilon, q]$, и разложим метрику в $q_c(\eta)$ в виде

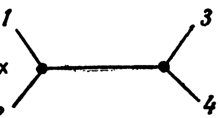
$$h_c^{\mu\nu} = g_c^{\mu\nu} - \gamma^{\mu\nu} = h_0^{\mu\nu} + \varepsilon h_1^{\mu\nu} + \dots,$$

где

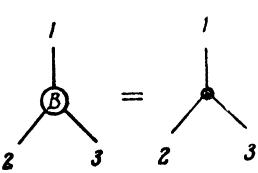
$$h_0^{\mu\nu} = \eta_1 f^{\mu\nu}(1) + \eta_2 f^{\mu\nu}(2) + \dots + \eta_N f^{\mu\nu}(N),$$

а $f(j)$ — решение линеаризованных уравнений поля, представляющее j-ю гравитационную волну в асимптотической области. Классическое возмущение $B(\eta)$ является суммой таких членов, как

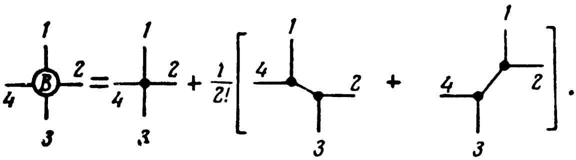
$$\eta_1\eta_2\eta_3\eta_4\varepsilon^2 f^i(1)\, I_{,ijk} f^k(2)\, G_0^{jr} f^s(3)\, I_{,srp} f^p(4) =$$

$= (\eta_1\eta_2\eta_3\eta_4) \times$ 

которые изображаются диаграммами согласно обычным правилам, где каждый множитель f^i представляется внешней линией, проведенной из вершин $I_{,ijk}$, с которой он свертывается, а каждый множитель G_0^{ij} — внешней линией, соединяющей две вершины, с которыми он свертывается. Все диаграммы, образованные разложением $B(\eta)$, являются деревьями. (Определяющей характеристикой дерева является то, что оно распадается на несвязанные части при разрезе любой линии.) Две примитивные скобки Фейнмана низшего порядка имеют вид



и



Обобщение диаграммных правил перестает быть таким непосредственным, когда учитываются радиационные поправки — диаграммы с замкнутыми петлями пропагаторов. Фейнман не нашел ясного пути для расчета этих процессов, сохранения калибровочной инвариантности и получения унитарной матрицы рассеяния. Однако он нашел способ получить унитарную калибровочно инвариантную матрицу рассеяния для гравитонов. Его основным инструментом были фейнмановские скобки. Суммируя примитивные скобки одинакового типа, но с различным выбором внешних функций $f^i(r)$, можно построить калибровочно инвариантные скобки, где вершины связаны двухточечной функцией Δ^{ij}, которая находится «на массовой оболочке» в том смысле, что она удовлетворяет линеаризованным уравнениям Эйнштейна и дополнительным условиям. Фейнмановский подход к расчету диаграмм с петлями заключается в замене их диаграммами, в которых по крайней мере один множитель G_0^{ij} в каждой петле заменяется множителем Δ^{ij}. Процесс описывается как «разрыв петель пропагаторов» и в значительной степени определяется требованием, чтобы получаемые фейнмановские скобки были полными, и унитарностью получаемой S-матрицы.

Де Витт [25, стр. 773, 774] сформулировал правила замены для «разорванных петель» следующим образом. Сначала разлагаем каждый фейнмановский пропагатор G_0^{ij} на опережающую или запаздывающую функцию Грина плюс двухточечную функцию на массовой оболочке. Это разложение можно осуществить несколькими способами, но только один из них правильный. Чтобы фиксировать правильную процедуру, заметим, что некоторые из диаграмм в сумме, получающейся в результате разложения, содержат *непричинные петли* — опережающие и запаздывающие функции Грина, связывающие голову и хвост диаграммы. Выбираем разложение, дающее максимальное число диаграмм с непричинными петлями. Во-вторых, отбрасываем все эти непричинные диаграммы. Оставшиеся диаграммы не должны содержать петель, полностью образованных функциями Грина. Каждая диаграмма является частью скобки Фейнмана.

В-третьих, следует добавить некоторые диаграммы, чтобы сделать полными все скобки Фейнмана и таким образом обеспечить калибровочную инвариантность. Наконец, чтобы разрешить любые остающиеся неоднозначности, проверяем унитарность полученной S-матрицы.

Фейнман [38] заметил, что правила замены для разрыва петель можно получить из обычных фейнмановских правил, если ввести фиктивные частицы. Эти фиктивные частицы представляют собой векторные фермионы, которые появляются только в промежуточных состояниях. Де Витт [22, стр. 780], а также Фадеев и Попов [33] показали, что фиктивные частицы соответствуют выбору более общего интегрального функционала, чем функционал, дающий обычные правила Фейнмана (аддитивная однородная группа, весовой множитель 1). Если $\int\limits_C \mathscr{D}g$ есть сумма по $g^{\mu\nu}$, удовлетворяющая координатным условиям C и инвариантная относительно аддитивных трансляций $g^{\mu\nu} \to g^{\mu\nu} + a^{\mu\nu}$, то правильная сумма по пространственно-временным историям имеет вид

$$\int \mathscr{D}q \ldots = \int\limits_C \mathscr{D}g \Delta(q, q_0) \ldots .$$

Плотность $\Delta(q, q_0)$ выражается в форме интегрального функционала

$$\Delta(q, q_0) = \int \mathscr{D}A \left\{ \exp\left(\frac{1}{2} i \int \sigma g^{\mu\nu} A^{\alpha *}_{\cdot\mu} A^{\beta}_{\cdot\nu} \gamma_{\alpha\beta} \right) \right\},$$

где $\int \mathscr{D}A$ аддитивно по A, A^{μ} интерпретируются как антикоммутирующие c-числа, а $A(\pm\infty)$ фиксируется равным нулю. Фадеев и Попов показали, что это выражение для плотности можно получить, если предположить, что $\int \mathscr{D}q$ является суммой по всем $g^{\mu\nu}$ и существует аддитивная однородная группа в $g^{\mu\nu}$. Тогда каждое различное пространство-время будет представлено набором $g^{\mu\nu}$, отличающимся друг от друга только выбором координат. Такой набор называют *орбитой* общековариантной группы (см. § 8). Тогда сумма $\int\limits_C \mathscr{D}g$ по различным пространствам-временам вычисляется с учетом весового множителя, равного «объему» орбиты, соответствующему каждому пространству-времени. Фактически этим объемом служит $\Delta(q, q_0)$.

Весовой функционал Δ и фиктивное поле A, которое его образует, появляются благодаря допущению, что однородная группа является абелевой, тогда как в случае группы координатных преобразований это не так. Фейнман [38] отметил, что аналогичная ситуация возникает из неабелевой калибровочной группы безмассового поля Янга — Миллса. Имеется другой путь обнаружить эту ситуацию, если заметить, что при изменении опорной геометрии q_0 сумма $\int\limits_C \mathscr{D}g$ будет изменяться сложным образом. Если сумма по пространствам-временам не зависит от q_0, то *должен* существовать компенсирующий весовой функционал. Но $\Delta(q, q_0)$ должна преобразовываться как функциональная *плотность* так, чтобы $\int \mathscr{D}q = \int\limits_C \mathscr{D}g$ было инвариантным относительно произвольных преобразований переменных интегрирования. Де Витт [22, стр. 774—781] заметил, что приближением наинизшего порядка для $\Delta(q, q_0)$ является простейшая возможная плотность, которую можно построить только из функционала действия, а именно

$$\Delta(q, q_0) = [\det I_{,ij}]^{1/2}.$$

Появление фиктивных частиц дает нежелательный эффект ослабления локальности и причинности в теории поля. Плотность $\Delta(q, q_0)$ определенно имеет нелокальное, непричинное происхождение. Сами фиктивные частицы нарушают причинность вследствие требования их исчезновения из наблюдаемых состояний. Появление нелокальности и непричинности относительно *произвольного* опорного многообразия q_0 не обязательно является нежелательным. Однако мы видели, что оно является неизбежным следствием калибровочно инвариантного квантования. Чтобы получить явно локальную и причинную теорию, следует отказаться от очевидной калибровочной инвариантности и выбрать предпочтительное опорное пространство-время (например, пространство-время Минковского) и предпочтительные дополнительные условия.

4. Операторы и временное упорядочение

Квантовую ОТО желательно привести к операторной форме, поскольку каноническая операторная квантовая теория поля привычна, хорошо исследована и операторный формализм представляет собой традиционные рамки, в которых развивалась теория поля. Формализм векторов-историй, которым мы пользовались, лишь частично удовлетворяет этим условиям. Чтобы получить полный операторный формализм, следует выразить операторы $T(\sigma\pm)[Q(\phi)]$ в виде сумм произведений полевых операторов $T(\sigma\pm)[\phi] = \Phi$. Как мы видели при рассмотрении обобщенных функций Грина, основное правило состоит в отождествлении $T(\sigma\pm)[Q(\phi)]$ с хронологическим произведением $T[Q(\Phi)]$. В лоренц-инвариантной теории $T[\ldots]$ не зависит от семейства гиперповерхностей σ. Однако это замечание подразумевает, что $T(\sigma\pm)[\ldots]$ не обязательно независимо от σ. Для простых теорий поля без калибровочных групп на это предупреждение можно не обращать внимания. Для абелевых калибровочных полей (квантовая электродинамика) хронологическое произведение необходимо дополнить швингеровскими членами, зависящими от σ [14, 52]. В ОТО калибровочная группа неабелева и фактически отождествляется с выбором пространственноподобных гиперповерхностей. Таким образом, имеются основания сомневаться в применимости обычного хронологического правила. Действительно, поскольку пространственно-временная метрика сама является динамическим полем, возникают сомнения в том, что хронологическое произведение можно определить в обычном смысле.

Мандельстам [65] разработал последовательную и вполне общую процедуру получения операторных формализмов для калибровочных полей. Основными составными частями этой процедуры является множество калибровочно зависимых объектов отсчета, которые можно скомбинировать с локальным калибровочно зависимым полем, чтобы образовать полный набор калибровочных инвариантов, множество стандартных объектов отсчета, локализованных или в асимптотической области, или в некоторой стандартной точке пространства-времени, правила переноса объектов отсчета вдоль путей в пространстве-времени и набор путей, связывающих каждую пространственно-временную точку с полным набором стандартных объектов отсчета. В ОТО объектами отсчета являются тетрады ортогональных векторов, которые должны параллельно переноситься вдоль данных путей. «Калибровочно инвариантными» объектами тогда будут тетрадные компоненты тензора кривизны. Конечно, инварианты, построенные таким образом, зависят от пути, связывающего их со стандартными тетрадами отсчета. Однако разность между тетрадами, перенесенными вдоль двух слегка различных путей в ту же самую точку пространства-времени, равна как раз параллельному переносу тетрады вдоль петли, образованной двумя путями. Таким образом, эту разность можно подсчитать при помощи тензора кривизны вдоль пути. Таким способом можно построить путезависимые соотношения, связывающие инфинитезимальные вариации путезависимых переменных в данной точке пространства-вермени с кривизной в других точках.

Подход Мандельстама не исключает проблемы калибровки, он просто переводит ее в более удобную и физически значимую форму, заменяя калибровочную зависимость зависимостью от пути. В путезависимом формализме можно определить пару пространственно-временных точек, которые *всегда* разделены пространственноподобным интервалом, и, таким образом, прийти к твердому отождествлению переменных, коммутирующих друг с другом. Понятие временного упорядочения можно, по крайней мере в принципе, определить в рамках особых путеопределенных процедур [64]. Путезависимые переменные являются именно теми переменными, которые измеряет реальный наблюдатель, движущийся на космическом корабле и использующий инерциальную навигацию. Таким образом, имеются серьезные основания для выбора этих переменных в качестве предпочтительных наблюдаемых.

Путезависимые компоненты тензора кривизны Римана [$R_{\alpha\beta\mu\nu}(x, P)$ в обозначениях Мандельстама, где P обозначает путь] рассматриваются как операторы с коммутационными соотношениями, даваемыми правилом коммутации. Хронологические произведения компонент кривизны относительно опорного пространства-времени (пространство-время Минковского) определяются обычным образом. Ковариантные хронологические произведения, характеризуемые своими простыми путезависимыми соотношениями, получаются прибавлением дополнительных членов к обычным хронологическим произведениям. Обобщенные функции Грина, полученные из ковариантных хронологических произведений, определяются путезависимыми соотношениями и наложением эйнштейновских уравнений поля в форме

$$\mathbf{R}_{\mu\nu\alpha\beta}\mathbf{g}^{\mu\alpha}=0$$

на каждый сомножитель произведения. Для решения определяющих уравнений Мандельстам ввел вспомогательные функции Грина, удовлетворяющие дополнительным условиям. Затем используются методы теории возмущений, чтобы выразить путезависимые функции Грина, а следовательно, элементы S-матрицы полностью через калибровочно зависимые вспомогательные функции Грина. Получаемое разложение тождественно разложению де Витта, включая члены, которые можно рассматривать как вклады фиктивных частиц.

Чтобы выразить процедуру Мандельстама на языке векторов-историй, который мы использовали выше, представим его ковариантное хронологическое произведение в виде

$$T(\sigma \pm)[R(x, P)\, R(x', P') \ldots].$$

Разложим компоненты кривизны по $h_{\mu\nu} = g_{\mu\nu} - \gamma_{\mu\nu}$ и получим путезависимые функции Грина, выраженные через функции Грина G_0^{ij}. Последние, конечно, являются мандельстамовскими калибровочно зависимыми функциями Грина. Для получения мандельстамовских путезависимых соотношений для путезависимых функций Грина потребуем, чтобы функциональное интегрирование было калибровочно независимым, что, как мы видели выше, в свою очередь требует наличия множителя плотности $\Delta(q, q_0)$, который вызывает появление в S-матрице членов, соответствующих фиктивным частицам.

5. Расходимости

Одним из основных мотивов квантования ОТО является надежда на то, что гравитационное взаимодействие может устранить расходимости электромагнитных и других теорий. Конечно, надеются также на то, что сама квантовая ОТО не слишком расходящаяся; в противном случае гравитационное взаимодействие просто перенесло бы расходимости в другое место без их ликвидации — знакомое явление в квантовой теории поля (например, [62]). В настоящее время появились свидетельства, что эти надежды оправданы. Главная трудность связана с ультрафиолетовой расходимостью. Инфракрасная расходимость гораздо менее серьезна, так как обычно ее можно устранить с помощью подходящего редукционного формализма [44]. Тем не менее стоит отметить, что в квантовой ОТО нет инфракрасной расходимости [26, 89].

На каждом этапе решения уравнений (6.9) относительно функций Грина с помощью метода итераций сталкиваются с неопределенными объектами, такими, как $\delta^2(x - x')$. Чтобы избежать этой трудности, вводят регуляризацию, или процедуру обрезания. Для этого имеется много способов. С той точки зрения, которой мы придерживаемся, наиболее естественной и хорошо определенной процедурой является замена уравнений (6.9) последовательностью систем уравнений, дающих в пределе уравнения (6.9). Каждая приближенная система характеризуется массой обрезания Λ. Вычитание Λ-приближенных уравнений из (6.9) дает сумму остаточных членов, являющихся кратными интегралами. Каждое подынтегральное выражение содержит произведение «пробных функций» $\delta g^{\mu\nu}(P)$. Мы требуем, чтобы остаточные члены стремились к нулю, как стремится к нулю наибольшее значение $\Lambda^{-1}\delta g^{\mu\nu}{}_{,\alpha}$. Затем, используя метод итераций, находим обобщенные функции Грина, являющиеся решениями Λ-приближенной системы. Тогда S-матрица разлагается в ряд по Λ-приближенным функциям Грина. Далее, чтобы получить правильную S-матрицу, устремляем Λ к бесконечности. К сожалению, предел не существует, или, более точно, не существует пределов для отдельных членов, входящих в разложение S-матрицы. В этом состоит существо проблемы ультрафиолетовой расходимости.

Общепринятая интерпретация проблемы ультрафиолетовой расходимости состоит в том, что функционал действия, приводящий к уравнениям (6.9), некорректен. Чтобы сделать S-матрицу сходящейся, к «голому» действию добавляют Λ-зависимые контрчлены. К сожалению, каждый контрчлен действует не только на расходящиеся части S-матрицы, но и на сходящиеся. В результате каждый контрчлен вносит элемент произвольности (обычно одну или более произвольных констант) в окончательную S-матрицу. Если все расходимости можно устранить введением конечного числа контрчленов, так что окончательная S-матрица содержит конечное число произвольных констант, то теория называется перенормируемой. Произвольные константы можно рассматривать как величины, перенормированные значения которых фиксируются экспериментом.

Расходящийся функционал действия, появляющийся из перенормировки, не влияет на какую-либо наблюдаемую величину, однако формально его появление весьма неприятно. Возникает подозрение, что расходимости перенормируемых теорий вытекают из неподходящего формализма, а не из глубоких недостатков квантовой теории поля. Таким образом, может оказаться, что в квантовой ОТО нет необходимости устранять расходимости других полей и она сама может не быть свободной от них. Если квантовая ОТО просто перенормируема и делает другие теории перенормируемыми, то это спасет квантовую теорию от проблем расходимости.

§ 7. КВАНТОВАНИЕ ЗАМКНУТЫХ МОДЕЛЕЙ ВСЕЛЕННОЙ: НОВЫЙ ТИП ФИЗИКИ

Замкнутые модели вселенной всегда занимали видное место в классической ОТО. Хотя астрономические доказательства еще не являются определенными, весьма вероятно, что мы живем в замкнутой вселенной. Далее, наша вселенная, по-видимому, начала существование с сингулярности («большого взрыва»), когда важную роль могли играть квантовые эффекты. Таким образом, есть основания изучать квантовую ОТО замкнутых моделей вселенной. В идеальном случае открытые и замкнутые вселенные следует рассматривать как возможные состояния в единой квантовой ОТО. Однако единой основы, которая реально позволила бы рассматривать вместе открытые и замкнутые вселенные, не существует. Чтобы добиться какого-либо прогресса, мы вынуждены игнорировать переходы между открытыми и замкнутыми вселенными. Для описания замкнутых вселенных мы будем использовать тот же язык векторов-историй, который мы применяли для описания квантования открытых вселенных. Однако это единство описания не изменяет того факта, что открытые и замкнутые вселенные трактуются как фундаментально различные физические системы.

Материал этого параграфа разделяется на две общие категории: 1) материал, который не ограничивается замкнутыми вселенными, но является полезным только тогда, когда неприменимы методы квантования открытых вселенных, и 2) материал, применимый только к замкнутым вселенным. Так, мы рассматриваем проблему замороженного формализма, который универсален, но становится действительно серьезным лишь в замкнутых вселенных, «гипотезу суперпространства», которое существенно ограничивается замкнутыми вселенными, квантование Гамильтона — Якоби, справедливое в общем случае, но ценное пока лишь в замкнутых вселенных, и различные модельные теории, специально предназначенные для представления замкнутых вселенных.

1. Проблема замороженного формализма в общековариантных теориях

В общековариантной квантовой ОТО *пространственно-временные* компоненты метрики рассматриваются как независимые друг от друга. Таким образом, включающее интегральный функционал выражение (5.33) для скалярного произведения можно интерпретировать как сумму по различным пространствам-временам, входящим с усреднением по всем пространственно-временным координатным системам (см. [33, 54], а также § 6, п. 1, настоящей статьи). Предполагается, что результат объединенного суммирования и усреднения инвариантен относительно пространственно-временных преобразований координат. Было показано (см., например, [66]), что эта пространственно-временная картина прямо приводит к характерной особенности общековариантных теорий: проблеме замороженного формализма.

На языке векторов-историй легко получить замороженный формализм, но он становится «проблемой» только в том случае, если настаивать на дополнительных ограничивающих предположениях. Обозначим координатную вариацию $g_{\mu\nu}$ через δ_c и, используя уравнение (5.7), получим

$$\langle F \mid \delta_c G \rangle = -i \langle F \mid \delta_c I \mid G \rangle = 0 \quad \text{для всех} \quad F \quad \text{и} \quad G. \tag{7.1}$$

Этот результат и называется замороженным формализмом. Он говорит, что любые два вектора-истории, которые можно связать координатным преобразованием, представляют одно и то же физическое состояние. Если положить δ_c вариацией временно́й координаты $t \to t + \delta t$, то уравнение (7.1) будет представлять собой в точности уравнение Шредингера, возникающее из обращения в нуль канонического гамильтониана [см. уравнение (5.19) и его обсуждение].

При использовании этого формализма векторов-историй не сразу видно, что уравнение (7.1) ставит какую-либо проблему. Ясно, что векторы-истории F и G не обязательно являются *функционалами состояний* в данном *координатном* времени. Проблема возникает, когда настаивают на том, что F и G *являются* функционалами состояний в данном координатном времени. Затем приходится считать F и G функционалами из полного набора сохраняющихся величин, которые вычисляются на пространственноподобной гиперповерхности, но не зависят от выбора этой гиперповерхности. Этот подход, использующий сохраняющиеся величины, исследовали Бергман [11] и Комар [58, 59].

Можно представить себе действительно серьезную проблему типа замороженного формализма, которая, однако, не следует из уравнения (7.1). Такая проблема возникла бы, если бы оказалось, что квантовая ОТО представляется только одним допустимым вектором-историей. Ключевой вопрос здесь заключен в смысле, вкладываемом в понятие «допустимый». Если все векторы-истории допустимы, то замороженного формализма не может быть. Однако *могут* быть «странные» представления. Если попытаться наложить слишком много ограничений, то, как мы видели в § 5, п. 3, можно исключить все представления, кроме одного, и прийти к истинно замороженному формализму.

В открытых вселенных проблема истинно замороженного формализма не возникает. Используя асимптотическую группу симметрии, можно построить по крайней мере два различных физически допустимых представления — входящее и выходящее.

В замкнутых вселенных, по-видимому, нет естественной группы, которую можно использовать для определения допустимых состояний, и нельзя исключить возможность того, что в действительности имеется только одно истинное допустимое представление.

2. Суперпространственные представления общековариантного квантования замкнутой вселенной

Квантовая ОТО Уилера — де Витта была введена в § 5, п. 1, как пример канонического квантования. Здесь мы используем ее, чтобы ввести часть гипотезы суперпространства. Эта гипотеза дает естественные ответы на вопросы, возникающие в результате попыток получить каноническую квантовую ОТО Уилера — де Витта на базе наших векторов-историй. Получение теории Уилера — де Витта на основе векторов-историй не приводит к законченной формулировке гипотезы суперпространства. Скорее оно показывает, почему такая гипотеза *необходима*, и обрисовывает ее *существенные* особенности.

Сделаем краткий обзор требований квантовой теории векторов-историй. Во-первых, требуются функционалы действия $I[q]$, определенные на историях q. Второе требование, очень важное для наших настоящих целей, — определение «представления квантовых состояний на основе векторов-историй». Кроме этого, необходимо знать, какие представления являются «странными» и какие допускают вероятностную интерпретацию с точки зрения экспериментов, которые в принципе можно осуществить. Третье требование: операции суммирования по историям $\int \mathcal{D}q \ldots$ должны характеризоваться однородными преобразованиями и, возможно, также мнимой поправкой к функционалу действия (см. § 6, п. 1).

Возьмем в качестве первичного функционала действия для квантовой ОТО действие АДМ, заданное уравнением (3.19). Эта форма действия объединяет каноническую форму с наибольшей возможной ковариантностью (т. е. пространственно-временной ковариантностью). Это действие зависит от времязависимого метрического тензора $g_{ij}(P, t)$ и времязависимых функций смещения и сдвига $N_\mu(P, t)$, поэтому мы предполагаем, что этих величин достаточно для описания истории q. Заметим, что мы не определяем еще термин «история» и не предполагаем, что описание с помощью g_{ij}, N_μ является наиболее общим или хотя бы наиболее подходящим.

Переходя к следующему требованию векторов-историй, мы спрашиваем: «Что такое представления квантовых состояний на основе векторов-историй?» или, более точно, «От каких свойств q может зависеть функционал по историям $F[q]$ в представлении $\mathfrak{r}$?» Более полезно вопрос поставить так: «Какие свойства q должны фиксироваться вариациями $\delta_{\mathfrak{z}}$ и $\delta_{\mathfrak{z}\times\mathfrak{z}'}$?» Чтобы избежать «странных» представлений, вариации должны фиксировать как можно меньше. С другой стороны, необходимо фиксировать достаточное число величин, чтобы уравнение (5.36) давало классические уравнения движения без остаточных поверхностных членов. Байерлейн и др. [7] показали, что для замкнутых вселенных существует единственное решение с этими условиями: $\delta_{\mathfrak{z}\times\mathfrak{z}'}$ точно фиксирует две *3-геометрии*. Они являются внутренними геометриями начальных и конечных пространственноподобных гиперповерхностей, служащих границами интеграла действия. Функции смещения и сдвига, пространственные координаты и временная координата *не должны* фиксироваться. Таким образом, допустимые представления, связанные функционалами действия АДМ, должны иметь форму

$$F[q]=\Psi(^3\mathcal{G}) \quad \text{для любого} \quad q \quad \text{и любого} \quad F \quad \text{в} \quad \mathfrak{r}, \qquad (7.2)$$

где $^3\mathcal{G}$ — пространственная геометрия на истории q.

Заметим, что эти соображения теряют силу в открытых вселенных, где есть дополнительные остаточные поверхностные члены в пространственной бесконечности. В открытой вселенной правильное представление должно включать корреляции между геометрией и внешними «лабораторными» приборами [24, 50].

Сравним уравнение (7.2) с соответствующим утверждением в механике частиц

$$F[q]=\psi(x, t) \quad \text{для любого} \quad q \quad \text{и для любого} \quad F \quad \text{в} \quad \mathfrak{r}(t). \qquad (7.3)$$

Из этого сравнения следует основное утверждение гипотезы суперпространства: *роль, которую в механике частиц играет пространство-время Минковского, в ОТО выполняет многообразие суперпространства, точками которого являются трехмерные геометрии.*

Прежде чем мы сможем использовать эту гипотезу, мы должны решить, имеется ли только одно допустимое представление вида (7.2) или много таких представлений. Наиболее естественное допущение и единственное, которое позволяет продолжить дальнейшее обсуждение на основе векторов-историй, состоит в том, что уравнению (7.2) удовлетворяют много представлений. При таком допущении считается, что некоторые свойства данной 3-геометрии определяют представление. Эти свойства, которые играют ту же роль, что и t в уравнении (7.3), мы обозначим ${}^3\mathcal{G}_L$, где L означает «продольные». Остальные свойства, необходимые для характеристики 3-геометрии, обозначим ${}^3\mathcal{G}_T$, где T означает «поперечные». Мы еще не знаем, как разделить свойства 3-геометрии на продольные и поперечные компоненты. Если продольные свойства должны играть роль временных переменных, их необходимо выбрать так, чтобы каждое возможное ${}^3\mathcal{G}_L$ находилось точно один раз в каждой возможной истории в суперпространстве. Следует ограничить дозволенные истории, чтобы запретить построение «плохих» историй, в которые некоторые ${}^3\mathcal{G}_L$ входят дважды или совсем не входят. Для таких ограниченных историй можно построить представления, удовлетворяющие уравнению (7.2), определив функционал $q:\Sigma({}^3\mathcal{G}_L)$, приписывающий истории q поперечные свойства той 3-геометрии в q, которая имеет продольные свойства ${}^3\mathcal{G}_L$. Этот функционал в точности аналогичен «смещению в данный момент времени» $q:\Sigma(a)$, которое мы рассматривали в § 5, п. 3. Тогда представление $\mathscr{r}[{}^3\mathcal{G}_L]$ состоит из функционалов-историй вида

$$F[q]=\Psi[q:\Sigma({}^3\mathcal{G}_L)]. \tag{7.4}$$

Выражение (7.4) уже содержит две основные черты теории Уилера — де Витта: функционал состояния не зависит от N_μ и координат, описывающих ${}^3\mathcal{G}$. Ранее мы вывели эти свойства на основе общековариантной канонической теории. Здесь мы пришли к ним в результате поисков допустимых представлений векторов-историй, связанных действием АДМ. Последнюю особенность теории Уилера — де Витта — уравнение де Витта (5.20) — можно получить многими способами.

Один способ получения уравнения (5.20) состоит в следующем: 1) берем g_{ij} и N_μ в качестве величин, описывающих историю q; 2) рассматриваем величину

$$\int_{\mathscr{r}({}^3\mathcal{G}_L)}^{\mathscr{r}({}^3\mathcal{G}'_L)} \mathscr{D}q\ \ldots,$$

которая объединяет суммирование по различным историям и усреднение по эквивалентным (g_{ij}, N_μ)-описаниям; 3) выбираем однородную группу с преобразованиями строго локального вида

$$g^\pi_{ij}(P)=\pi_{ij}(g_{rs}(P),\ P),$$

$$N^\pi_\mu(P)=\pi_\mu(N_\nu(P),\ P);$$

4) берем в качестве фазового функционала уравнения (5.33) действие АДМ, деленное на L^2 [см. обсуждение, следующее за уравнением (5.34)]; 5) используем тождество (5.37) для получения представлений операторов импульсов через вариационные производные; 6) используя только одну вариацию N в уравнении (5.36), получаем уравнение де Витта.

Простейший способ осуществить намеченный путь получения уравнения де Витта — выбрать дополнительные однородные преобразования

$$g^{\pi}_{ij}(P) = g_{ij}(P) + \pi_{ij}(P).$$

Однако поучительно выбрать преобразования

$$g^{\pi}_{\mu\nu} = \pi^{\alpha}_{\mu} g_{\alpha\beta}(P)\, \pi^{\beta}_{\nu},$$

предложенные Мизнером [66] и Клаудером [54]. Инфинитезимальная $(3+1)$-форма этих преобразований имеет вид

$$\delta N = N\,\delta a + (\vec{N}^2 - N^2)\left(\frac{N^i}{N}\right)\delta b_i + 2\left(\frac{N^i}{N}\right) N_j\, \delta a^j_i,$$

$$\delta N_i = N_i\, \delta a + g_{ij}\,\delta a^j + (\vec{N}^2 - N^2)\,\delta b_i + N_j\,\delta a^j_i,$$

$$\delta g_{ij} = N_i\,\delta b_j + g_{ik}\,\delta a^k_j + \text{выражение, симметричное по } i \text{ и } j,$$

где

$$\vec{N}^2 \equiv N_i N^i \equiv g_{ij} N^i N^j,$$

а δa, δa^j, δb_j и δa^k_j — произвольные инфинитезимальные пространственно-временные функции. Уравнение (5.37), в котором δa, δa^j и δb_j равны нулю, дает соотношение

$$T(\boldsymbol{r})\,[\boldsymbol{\pi}\mathbf{g}]\,|F\rangle = -\,iL^2 \left|\left(\frac{\delta F}{\delta \mathbf{g}}\right)\mathbf{g}\right\rangle, \tag{7.5}$$

где указаны матричные произведения. Пока F равно нулю для любой геометрии в точке P, в которой $g_{ij}(P)$ — сингулярная матрица, мы можем прямо заключить, что (5.13) дает представление для $T(\boldsymbol{r})\,[\boldsymbol{\pi}]$ посредством вариационной производной. Однако следует заметить, что непосредственно представимо только матричное произведение $\boldsymbol{\pi}\mathbf{g}$. Этот результат аналогичен результату, получающемуся из аффинных коммутационных соотношений Клаудера. *Преобразования Мизнера генерируют те же представления операторов, что и подход, использующий аффинные канонические операторы.* Полученное операторное представление отличается от стандартной канонической теории только в трактовке сингулярных метрик ($\det \mathbf{g} = 0$). Если эти сингулярности исключить граничным условием, то никакой разницы не будет. Получим теперь уравнение де Витта, положив, что δa является единственным ненулевым коэффициентом в вариации, входящей в уравнение (5.36). Вспоминая, что члены с $\delta I/\delta N^i$ исчезают вследствие уравнения (7.4) и что $\delta I/\delta N(P) = \mathscr{H}(P)$, находим

$$\langle G \,|\, T(t)\,[N(P)\,\mathscr{H}(P) \,|\, F\rangle = 0 \tag{7.6}$$

для всех G и F. Теперь внесем предположение об упорядочении сомножителей

$$T(t)\,[N(P)\,\mathscr{H}(P)] = \{T(t)\,[N(P)]\}\,\{T(t)\,[\mathscr{H}(P)]\}$$

и, заметив, что соотношение (7.4) препятствует тому, чтобы в результате действия $T(t)\,[N(P)]$ аннигилировало какое-либо допустимое состояние, найдем

$$T(t)\,[\mathscr{H}(P)]\,|\,F\rangle = 0.$$

Это соотношение вместе с операторными представлениями (7.5) и предположением о надлежащем упорядочении сомножителей приводит к уравнению де Витта (5.20).

Какими другими способами можно получить характерные особенности теории Уилера — де Витта? Существует несколько возможных путей другой интерпретации только что полученных доказательств. Например, пусть история описывается только g_{ij}; кроме того, заменим $\int \mathscr{D}q$ в приведенном выше доказательстве на $\int \mathscr{D}q \int \mathscr{D}n$, где $\int \mathscr{D}n$ — сумма-усреднение по всем N_μ. Затем рассмотрим

$$\int \mathscr{D}n \exp\left\{ \frac{i}{L^2} I_{\text{АДМ}}(g, n) \right\} \equiv \exp\{iI_{\text{RM}}(g)\}$$

как интегрально-функциональное представление I_{RM} — действия на многообразии $\text{Riem}(M)$, точки которого определены значениями g_{ij}, а далее будем действовать, как прежде. Дальнейшая интерпретация может выделить усреднение по 3-координатам, чтобы получить функциональное представление I_{SS} — функционала действия для однопараметрических кривых в суперпространстве. Еще дальше можно рассматривать истинную историю в суперпространстве как собрание, или «каталог» 3-геометрий, находящихся во взаимно однозначном соответствии с пространственноподобными сечениями пространства-времени. Затем следует попытаться выделить из уравнения (5.33) интегрально-функциональное представление действия на истинных суперпространственных историях. Все эти возможные интерпретации могут дать уравнение де Витта. Причина исследования этих возможных интерпретаций состоит в том, что мы вынуждены *отгадывать* однородную группу для $\int \mathscr{D}q$. Может существовать такая интерпретация, в которой правильная однородная группа очевидна.

Другой способ вывода основных особенностей теории Уилера — де Витта состоит в использовании известного канонического формализма АДМ. При этом степени свободы, описываемые функциями смещения и сдвига, исключаются путем решения уравнений связи до квантования. Пространственно-временные условия исключают дальнейшие ложные степени свободы. Оставшиеся степени свободы ($S^1, S^2, \ldots, S^\infty$) являются именно теми, которые определяют внутреннюю 3-геометрию гиперповерхности постоянного времени. В этом подходе переменные $S^1, S^2, \ldots$ появляются как времязависимые параметры в g_{ij}. На данном этапе еще существуют лишние степени свободы, соответствующие выбору гиперповерхностей постоянного времени в данном пространстве-времени. Для их исключения наложим бесконечный набор «рассекающих условий»

$$f_A(S, t) = 0.$$

Для заданного t эти условия фиксируют гиперповерхность $\Sigma(t)$ в любом данном пространстве-времени. Эти условия исключают некоторые S^A. Остающиеся степени свободы могут быть описаны «поперечными параметрами» S_T^A. На историях, удовлетворяющих всем связям и условиям, которые мы описали, функционал действия АДМ принимает известную гамильтонову форму

$$I = \int dt \left\{ \sum_A \dot{S}_T^A p_A - H_f(S, p) \right\}.$$

Индекс f при H означает зависимость этой формы от использованного рассекающего условия. Из этой гамильтоновой формы можно вывести обычную квантовую теорию с уравнением Шредингера

$$\pm \frac{\partial \psi_f}{\partial t_f} = i H_f \psi_f \tag{7.7}$$

для каждого набора рассекающих условий. Здесь ψ_f зависит только от t_f и S_T, определенного через f.

Связь уравнения (7.7) с уравнением де Витта затемняется вопросами упорядочения сомножителей. Однако эту связь можно обрисовать в общих чертах. Состояние ψ_f АДМ определено на подпространстве суперпространства, в котором разрешена только одна продольная степень свободы, а именно t_f. Таким образом, можно искать «главное состояние» $\Psi({}^3\mathcal{G})$, определенное во всем суперпространстве, которое дает ψ_f, когда оно вычисляется на подпространстве, определенном f. Если утверждать, что Ψ должно содержать и положительные и отрицательные частотные компоненты, то определяющее уравнение

$$\left(\frac{\partial\Psi}{\partial t_f}\right)^2 + H_f^2\Psi = 0 \quad \text{для всех} \quad f \tag{7.8}$$

имеет ту же структуру, что и уравнение де Витта (5.20), — бесконечный набор дифференциальных уравнений второго порядка для функционалов, определенных только на 3-геометрии. Если подходящим образом переопределить пространственно-координатные степени свободы, то можно из уравнения (7.8) получить уравнение (5.20).

3. Квантование Гамильтона — Якоби

В такой претенциозной программе, как квантование ОТО, не лишне рассмотреть промежуточную ступень между эйнштейновской классической теорией и полной квантовой теорией. Известное квазиклассическое приближение предполагает, что фаза функционала состояния изменяется очень быстро по сравнению с его комплексным модулем. При этом допущении комплексная фаза функционала состояния является решением классического уравнения Гамильтона — Якоби. Основное достоинство этого подхода заключается в том, что он рассматривает те аспекты квантовой теории, которые не зависят от проблем выбора мер или упорядочения сомножителей.

Внося квазиклассические предположения в уравнение де Витта (5.20), находим, что комплексная фаза S удовлетворяет уравнению

$$R + g^{-1}\left(\frac{1}{2} g_{ij} g_{rs} - g_{ir} g_{js}\right) \frac{\delta S}{\delta g_{ij}} \frac{\delta S}{\delta g_{rs}} = 0.$$

Это уравнение Гамильтона — Якоби ОТО [72]. Строго говоря, это не одно уравнение, а бесконечное число уравнений: каждой пространственной точке соответствует одно уравнение, так как в каждой пространственной точке существует независимый временной параметр. Удобное математическое описание этой структуры дается многовременным формализмом Томонага [83, 88]. Сначала вспомним, что траектория обычной динамической системы представляется кривой в ее конфигурационном пространстве. Для каждой точки этой кривой имеется определенное значение «временного» параметра. Как показал Томонага, теория поля допускает более общий тип траекторий. Вместо реальной линии рассмотрим четырехмерную параметрическую систему. Элементы траектории связаны с гиперповерхностями в пространстве параметров. На этом языке траектория уже не является кривой. Вместо этого получается каталог полевых (или геометрических) конфигураций, который возникает из всевозможных способов построения гиперповерхностей равного времени.

Важное приложение метода Гамильтона — Якоби состоит в получении уравнений Эйнштейна как классического предела квантовой ОТО Уилера — де Витта [41]. Основная идея состоит в построении классического волнового пакета из совокупности фазовых функционалов Гамильтона — Якоби. Пусть $S(\alpha)$ — фазовый функционал Гамильтона — Якоби, который соответствует постоянным интегрирования, обозначенным через α. Тогда условие, что функционалы состояния $A(\alpha, \mathscr{G}) \exp\{S(\alpha, \mathscr{G})\}$ должны все интерферировать в данной 3-геометрии $\mathscr{G}$, имеет вид

$$\frac{\delta S(\alpha, \mathscr{G})}{\delta \alpha} = 0.$$

Герлах [41], используя уравнения Гамильтона — Якоби для S и координатную независимость S, смог показать, что 3-геометрии, определенные этими условиями интерференции, должны принадлежать классическому каталогу. Они являются внутренними геометриями всех пространственноподобных гиперповерхностей, которые можно построить в эйнштейновском пространстве-времени. Таким образом, уравнение де Витта содержит все уравнения Эйнштейна в классическом пределе.

Когда мы в первый раз вводили уравнение де Витта (5.20), мы дали правдоподобную аргументацию для насыщения динамики ОТО гамильтоновыми уравнениями связи. Эта аргументация выражалась через ожидаемое значение эйнштейновского тензора $G_{\mu\nu}$. Однако этот тип доказательства не может быть убедительным, так как ожидаемые значения чувствительны к выбору предпочтительных канонических переменных или, что эквивалентно, к выбору меры на историях. Тем не менее вывод, использующий метод Гамильтона — Якоби, полностью независим от выбора меры или предпочтительных переменных.

4. Модельные теории

ОТО — бесконечномерная теория. Для решения даже классических уравнений ОТО необходимо исключить большинство из этих измерений. Успешная классическая процедура заключается в наложении пространственных симметрий и специальных алгебраических свойств. Например, сферически симметричный мир можно описать только одной переменной — его радиусом. Благодаря свойству классических уравнений сохранять такие симметрии можно не учитывать все переменные, описывающие отклонение от симметрии, и все же получить точные решения. По существу все точные решения уравнений Эйнштейна были получены этим способом. Однако в квантовой ОТО точные решения нельзя получить, не учитывая переменные, нарушающие пространственные симметрии. Каждый вид нарушения симметрии имеет флуктуации около нулевой точки, которые дают конечные вероятности для несимметричных пространств. По этой причине квантовые теории симметричных миров представляют собой *модельные* теории, которые являются лишь приближениями к полной теории и, возможно, в какой-то степени отражают ее общую структуру, но в общем не дают точных решений полной теории. Единственный простой способ выделить точную конечномерную специализацию квантовой ОТО заключается в наложении суперпространственных симметрий на квантовый функционал состояния. К сожалению, не известно, обладает ли само суперпространство какими-либо динамически важными симметриями [87]. Таким образом, по крайней мере в настоящее время, только различные *модели* квантовой ОТО дают какую-то возможность проводить детальные вычисления.

Можно различать три типа модельных теорий. Первый тип — *технические модели*, такие, как безмассовая теория Янга — Миллса [43] и модели Клаудера [6, 57], сформулированы для подражания одной или более техническим особенностям полной теории (неабелева калибровочная группа, положительная определенность канонических переменных), но не стараются приблизиться к ней прямо. Второй тип — *симметричные модели*, такие как квантованная фридмановская вселенная де Витта [24] и миксмастеровская вселенная Мизнера [67], стараются приблизиться к полной теории, а также и подражать ее техническим проблемам, игнорируя все детали, нарушающие данную пространственную симметрию. Наконец, последняя модельная теория, которую мы будем называть *неограниченной усеченной схемой*, состоит из *последовательности* модельных теорий возрастающей размерности, которая как угодно близко приближается к полной теории в соответствующем пределе. Мы рассмотрим достаточно подробно симметричные модели, предложенные Мизнером и де Виттом. Неограниченной усеченной схемы еще не существует. Однако это важная цель будущих исследований. Таким образом, мы обсудим некоторые ее возможные проблемы. Мы предпочитаем не рассматривать технические модели более подробно, чем это было сделано в § 6, где говорилось о поле Янга — Миллса и о проблемах неабелевой калибровки, и в § 5, п. 1, где кратко говорилось о проблемах положительной определенности.

Мизнер [67] предложил несколько пространственно-симметричных моделей квантовой ОТО. Эти модели связаны с однородными вселенными. Мизнер рассматривает квантовую теорию нескольких таких вселенных. Мы опишем анализ замкнутых однородных вселенных, в которых пространственные сечения топологически являются 3-сферами. Эти вселенные однородны относительно трехмерной группы $SU(3)$. Пространственно-временную метрику для геометрии этого типа всегда можно записать в виде

$$ds^2 = -N^2\,dt^2 + \frac{1}{4}R^2(t)\,[\exp\{2\beta(t)\}]_{ij}\,\sigma_i\sigma_j, \tag{7.9}$$

где

$$\sigma_1 = \sin\psi\,d\theta - \cos\psi\sin\theta\,d\phi,$$
$$\sigma_2 = \cos\psi\,d\theta + \sin\psi\sin\theta\,d\phi,$$
$$\sigma_3 = -(d\psi + \cos\theta\,d\phi).$$

Здесь величины N, R и β_{ij} зависят только от временной координаты t. Матрица β_{ij} имеет нулевой след, а $\exp\beta$ разлагается в ряд по матрице β. Матрица β измеряет анизотропию мира. Когда β обращается в нуль, мы получаем стандартную космологию Робертсона — Уолкера, которая однородна и изотропна.

В уравнении (7.9) величину R удобно заменить величиной Ω, определенной выражением

$$R = \left(\frac{2}{3\pi}\right)^{1/2}\exp(-\Omega).$$

Используя Ω и компоненты матрицы β в качестве динамических переменных, можно избежать проблем положительной определенности, рассмотренных в § 5, п. 1.

Ненулевые компоненты матрицы β имеют вид

$$\begin{aligned}\beta_{11} &= \beta_+ + 3^{1/2}\beta_-,\\ \beta_{22} &= \beta_+ - 3^{1/2}\beta_-,\\ \beta_{33} &= -2\beta_+.\end{aligned} \tag{7.10}$$

Без потери общности можно считать матрицу β диагональной. Тогда переменные β_+ и β_- измеряют инвариантный характер анизотропии. Чтобы понять, каким образом β_+ и β_- характеризуют анизотропию, полезно рассмотреть их как компоненты 2-вектора. Длина этого вектора измеряет общую суммарную анизотропию. Угол $\varkappa$, который вектор составляет с осью β_+, измеряет отклонение от аксиальной изотропии. Когда этот угол равен нулю или кратен $2\pi/3$, в каждой точке пространства существует ось вращательной симметрии, как во вселенной Тауба. Наибольшее отклонение от вселенной Тауба получается, когда $\varkappa$ равно $2\pi/6$ или $2\pi/6$ плюс любое число, кратное $2\pi/3$.

Переменные Ω, β_+ и β_- характеризуют 3-геометрии — они являются координатами на части суперпространства. Из нашего обсуждения анизотропии ясно, что все рассмотренные здесь вселенные можно представить внутри любого 60°-го клина, исходящего из начала плоскости β. Переход от одного такого клина к другому соответствует переименованию пространственных координатных осей, которые используются для определения β_+ и β_-. В суперпространстве, в котором каждой точке однозначно соответствует 3-геометрия, следует ожидать, что β_+ и β_- ограничены данным клином. Однако такого ограничения не делается, так как возникающие при этом «стенки» в суперпространстве вызвали бы недифференцируемые скачки для некоторых классических траекторий. Чтобы траектории в суперпространстве были настолько регулярными, насколько это возможно, на β_+ и β_- не должно накладываться никаких ограничений. Таким образом, мы находим, что «динамически естественное» суперпространство геометрий типа IX Бианки представляет каждую геометрию шесть раз — по одному разу в каждом 60°-ном клине плоскости β.

Используя гамильтонов формализм, развитый АДМ (см. § 4, п. 2), можно построить гамильтонов функционал действия для вселенных, описываемых уравнением (7.9) [67]. Это действие имеет вид

$$I = \int (p_+ \, d\beta_+ + p_- \, d\beta_- - H \, d\Omega), \tag{7.11}$$

где

$$H(p, \beta) = \{p_+^2 + p_-^2 + e^{-1\Omega}(V - 1)\}^{1/2} \tag{7.12}$$

и

$$V(\beta) = \frac{1}{3} \operatorname{Sp}(e^{4\beta} - 2e^{-2\beta} + 1).$$

В этой однородной модели, используя уравнение (7.9), можно разрешить пространственноподобные уравнения связей, положив

$N_i = 0$. Четвертую, или гамильтонову, связь необходимо разрешить относительно $N(t)$. Однако этот результат для функции сдвига $N(t)$ не играет непосредственной роли в динамике. Он влияет только на интерпретацию решений. Структура принципа действия предполагает, что β_+ и β_- можно взять в качестве канонических переменных с сопряженными импульсами p_+ и p_-, а также что Ω можно рассматривать как временную переменную с соответствующим гамильтонианом H.

Перепишем метрику (7.9) с $t = \Omega$ и подставим выражение для $N(t)$, являющееся решением гамильтоновой связи. В результате получим пространственно-временную метрику

$$ds^2 = -\left(\frac{2}{3\pi}\right) H^{-2} e^{-6\Omega}\, d\Omega^2 + (6\pi)^{-1}\, e^{-2\Omega} \left(e^{2\beta}\right)_{ij} \sigma_i \sigma_j,$$

в которой Ω играет роль допустимой временной переменной в пространственно-временной геометрии. На языке Ω-времени начало вселенной («большой взрыв») соответствует $\Omega = +\infty$. Универсальное собственное время покоящихся наблюдателей во вселенной обозначается τ; оно связано с Ω-временем соотношением

$$d\tau = N(\Omega)\, d\Omega = \pm \left(\frac{2}{3\pi}\right)^{1/2} H^{-1} e^{-3\Omega}\, d\Omega,$$

которое показывает, что вблизи $\Omega = \infty$ очень большие приращения Ω-времени вызывают довольно малые приращения τ-времени. Оказывается, что прошедшее с момента «большого взрыва» собственное время τ конечно, как в знакомой вселенной Фридмана.

Традиционная гамильтонова динамика, вытекающая из уравнения (7.11), полностью применима для анализа поведения вселенных в течение периодов равномерного расширения или сжатия. Однако для применения этого формализма к периодам перехода от расширения к сжатию или наоборот необходимо принять динамическое время, отличное от Ω, например универсальное собственное время τ. Но тогда теряется простое описание областей расширения и сжатия. Мизнер предложил способ получения динамики в простой и ясной форме как для фазы постоянного расширения или сжатия, так и для фазы перехода. Он использовал гамильтонов формализм, аналогичный формализму специальной теории относительности. Переменные β_+, β_- и $\beta_0 \equiv \Omega$ должны рассматриваться как координаты в трехмерном пространстве Минковского. Он определил ковариантный вектор импульса в данной точке вдоль траектории, взяв аффинный параметр λ вдоль траектории и положив

$$p_+ \equiv \frac{d\beta_+}{d\lambda}, \quad p_- \equiv \frac{d\beta_-}{d\lambda}, \quad p_0 \equiv \frac{d\beta_0}{d\lambda}.$$

Ковариантный функционал действия, дающий эти определения импульсов вместе с динамическими уравнениями, имеет вид

$$\bar{I} = \int d\lambda \left\{ p \cdot \frac{d\beta}{d\lambda} - K \right\},$$

где супергамильтониан

$$K = \frac{1}{2} \{ p \cdot p - e^{-4\Omega} (V - 1) \}.$$

Гамильтонов формализм, вытекающий из этого функционала действия, совершенно аналогичен общековариантному гамильтонову формализму, рассмотренному в § 4, п. 2. Полная общековариантная группа заменяется множеством преобразований аффинного параметра λ. Как и в общековариантном формализме, требуют обращения в нуль супергамильтониана K, чтобы гарантировать совместность уравнений с предположенными ковариантными свойствами. Таким образом, налагаем вторичное уравнение связи $K = 0$.

С точки зрения неквантовой или классической динамической задачи главное достоинство этого гамильтонова формализма специальной теории относительности состоит во вторичной связи $K = 0$. Используя только эту связь, можно выяснить, является ли траектория пространственноподобной, времениподобной или нулевой относительно выбранной метрики. Посмотрим, какие общие свойства классических траекторий можно получить таким способом. Различные теоремы, касающиеся неизбежности гравитационного коллапса (см., например, [48]), убеждают нас, что некоторая часть любой траектории лежит при $\Omega = \infty$. Из уравнения связи $K = 0$ совершенно ясно, что в этой области траектория должна быть нулевой или светоподобной относительно трехмерной метрики Минковского. Для анализа того, что происходит при конечных значениях Ω, т. е. после начального «большого взрыва» и до конечного коллапса, необходимо подробно исследовать структуру функции $V(\beta)$. Эта функция обращается в нуль только при $\beta_+ = \beta_- = 0$ и имеет поверхности уровня, которые являются кривыми, окружающими начало плоскости β. Поверхность $V = 1$ выделена и называется «узловой поверхностью». В суперпространстве (Ω, β) эта узловая поверхность является по существу треугольной призмой (с углами, отнесенными в бесконечность), окружающей ось Ω. Внутри узловой поверхности траектории должны быть пространственноподобными, т. е. $d\beta/d\lambda$ должно доминировать над $d\Omega/d\lambda$. Таким образом, пустые фридмановские вселенные, лежащие на оси Ω, запрещены. Однако фаза максимального расширения, для которой $d\Omega/d\lambda = 0$, может попасть в эту внутреннюю область. Вне узловой поверхности траектории должны быть времениподобными — $d\Omega/d\lambda$ должно доминировать над $d\beta/d\lambda$, так что фаза максимального расширения не может здесь происходить. Траектории, пересекающие саму узловую поверхность, должны двигаться в нулевом или светоподобном направлении в суперпространстве. Если принять, что аффинный параметр λ равен Ω (предполагая, что фаза максимального расширения не встречается), то получим $p_0 = 1$ и найдем, что уравнение связи $K = 0$ обеспечивает такое поведение траекторий, как если бы они двигались в потенциальной яме, где $V(\beta)$ — потенциал.

Чтобы завершить картину, заметим, что существуют необычные траектории, которые подходят ортогонально к узловой поверхности (при постоянном Ω и перпендикулярно поверхности в плоскости β). Такая траектория не достигает локальной «скорости света», требуемой для пересечения поверхности, и она просто кончается. Кроме того, такую траекторию нельзя согласовать с соответствующей траекторией, подходящей к поверхности с другой стороны, так как такие пространственноподобные траектории допускаются только внутри узловой поверхности. Траектории, оканчивающиеся на узловой поверхности, соответствуют симметричным по времени решениям уравнений Эйнштейна. Фаза расширения для решения соответствует траектории, следующей к узловой поверхности, в то время как фаза сжатия соответствует обратному движению в бесконечность.

Обсуждаемая теория является точной реализацией эйнштейновской неквантованной, или классической ОТО. Эту теорию можно проквантовать аналогично квантовой специальной теории относительности. Определим операторы импульса

$$p_0 = \frac{1}{i}\frac{\partial}{\partial\beta_0}, \quad p_+ = \frac{1}{i}\frac{\partial}{\partial\beta_+}, \quad p_- = \frac{1}{i}\frac{\partial}{\partial\beta_-}.$$

Тогда на функцию состояния $\Psi(\Omega, \beta_+, \beta_-)$ налагается уравнение «Клейна — Гордона» [68]

$$K_{op}\Psi = 0.$$

Сходство этого уравнения Клейна — Гордона с уравнением де Витта (5.20) очевидно как по структуре, так и по способу получения. Однако эти два уравнения вообще не эквивалентны, так как, во-первых, канонические переменные выбраны по-разному — с одной стороны, β_-, β_+, Ω, а с другой $g_{\mu\nu}(P)$, — и во-вторых, мизнеровское уравнение Клейна — Гордона выводится на основе специализированной теории, которая игнорирует или «замораживает» многие степени свободы. Таким образом, квантовая теория, полученная Мизнером, в лучшем случае является «модельной» теорией квантовой ОТО.

Для оценки справедливости модельной квантовой теории Мизнера необходимо исследовать значение замороженных степеней свободы. Они соответствуют более высоким модам вселенной или более коротким длинам волн гравитационного излучения. Как отметил Мизнер, Брилл и Хартл [19] показали, что очень коротковолновую гравитационную радиацию можно трактовать, как если бы это было электромагнитное излучение на фоне средней геометрии. Следовательно, игнорирование соответствующих мод не должно быть серьезной ошибкой. Трудная проблема связана с длинами волн, сравнимыми с радиусом вселенной, но достаточно короткими, чтобы нарушить однородность. Нелинейные пространственноподобные связи ОТО вызывают значительные трудности при попытке включить степени свободы, соответствующие этим волнам, в подход к квантованию АДМ.

Хотя исследования модельных квантовых теорий типа Мизнера находятся лишь в стадии разработки, во всех случаях, когда имеются точные решения или надежные приближения, по-видимому, проявляется одна характерная черта. Волновые пакеты в модельном суперпространстве расплываются не так, как волновые пакеты частиц в пространстве-времени. В настоящее время представляется, что неквантовые приближения ОТО с одинаковым успехом относятся как к ранней стадии «большого взрыва», так и к современной стадии. Это замечание находится в прямом противоречии с уже давно существующей надеждой, что квантование может спасти ОТО от трудностей, связанных с сингулярностями. Было бы интересно проследить за судьбой этого результата по мере исследования более изощренных модельных теорий и схем физической интерпретации.

Ряд трудностей мизнеровской модели и вообще любой попытки квантования ОТО возникает при стремлении дать физическую интерпретацию функционалу состояния Ψ. Можно, действуя так же, как в квантовой специальной теории относительности (см. [81], стр. 75), определить сохраняющийся благодаря уравнению Клейна — Гордона ток и назвать его током вероятности. Однако тот факт, что траектории в суперпространстве не обязательно остаются пространственноподобными или временниподобными, а даже могут быть замкнутыми во времени, слишком усложняет эту интерпретацию. Найдутся такие плотности вероятности, которые отрицательны в одном месте и положительны в другом. Решение этой очень трудной проблемы, вероятно, лучше всего оставить на будущее, когда будет лучше изучена динамика ОТО.

Модельная теория, которую мы рассматривали выше, началась как определенный тип традиционного гамильтонова формализма АДМ, а затем в какой-то степени вернулась к общековариантному формализму. Модель, с самого начала использующую общековариантное квантование, предложил де Витт [24]. Эта модель проще и однозначнее мизнеровской и более тесно касается проблем физической интерпретации и вопроса об определении временного параметра. Модель де Витта является вселенной Робертсона — Уокера [$\beta = 0$ в уравнении (7.9)]. Остаются только степени свободы, связанные с расширением и сжатием. Для сохранения сферической симметрии этой модели в классической теории требуется присутствие некоторого количества материи. Де Витт рассматривает эту материю как облако часов, назначение которого — обеспечить феноменологический временной параметр. Каждые часы имеют степень свободы q_i и лагранжиан $l(q_i, N^{-1}\dot{q}_i)$. Часы должны быть синхронизованы и распределены однородно и изотропно, так чтобы лагранжиан материи имел вид

$$L_m = MNl(q,\ N^{-1}\ \dot{q}),$$

где $q_i = q$ для всех i. Здесь M — суммарная масса всех часов, а N — функция сдвига, которая выбирается равной постоянной. Находим, что гравитационный лагранжиан этой ограниченной модели имеет вид

$$L_g = 12\pi^2(-N^{-1}R\dot{R}^2 + NR).$$

Пространственноподобные связи, как и в подходе Мизнера, автоматически удовлетворяются и на них можно не обращать внимания. Гамильтонова связь

$$\frac{\delta(L_g + L_m)}{\delta N} = \mathscr{H} = 0,$$

тогда дает уравнение

$$\mathscr{H}\Psi = 0.$$

Выражая $\mathscr{H}$ через R, q и сопряженные операторы

$$\pi = \frac{\delta L}{\delta \dot{R}} = \frac{1}{i}\frac{\partial}{\partial R}, \quad p = \frac{\delta L}{\delta \dot{q}} = \frac{1}{i}\frac{\partial}{\partial q},$$

получаем уравнение модели де Витта

$$\left\{\frac{1}{48\pi^2} R^{-1/4}\frac{\partial}{\partial R} R^{-1/2}\frac{\partial}{\partial R} R^{-1/4} - 12\pi^2 R + Mm\right\}\Psi(R, q) = 0,$$

где гамильтониан отдельных часов имеет вид

$$m = qp - l.$$

Это уравнение модели де Витта второго порядка по R. Таким образом, аналогично знакомому уравнению Клейна — Гордона оно имеет структуру начальных значений, которая не подходит для квантовой теории. Для уравнения Клейна — Гордона эта проблема решается наложением дополнительного граничного условия положительной определенности энергии, чем запрещаются отрицательно частотные решения. Де Витт доказал, что для уравнения модели де Витта единственным граничным условием должно быть равенство нулю Ψ при $R = 0$. Это граничное условие подразумевает, что сингулярная область («большой взрыв» или коллапс) в некотором смысле запрещена в квантовой теории. Одним из результатов этого предположения является то, что вселенная после коллапса «отскакивает» обратно, совершая периодические сжатия и расширения.

В одном отношении решения уравнения модели де Витта похожи на решения модели Мизнера: волновые пакеты не расплываются в плоскости (R, q). Однако ни R, ни q полностью не удовлетворяют требованиям к временной координате. Де Витт показал, что вместо q можно использовать феноменологический временной оператор τ, сопряженный с гамильтонианом часов. Можно видеть, что в τ-времени волновые пакеты расплываются вследствие возрастающих неопределенностей в значении самого τ.

Граничное условие де Витта при $R = 0$ равносильно внесению добавки к классической теории, которая решает проблему сингулярностей, предсказывая, что произойдет «после» конечного сжатия вселенной. Однако важность требования обращения в нуль Ψ при $R = 0$ совсем не ясна, так как интерпретация модели квантовой теории де Витта совсем не проще интерпретации модели Мизнера. Необходимо определить сохраняющийся ток, а затем решить, как избежать или интерпретировать отрицательные вероятности, появляющиеся из-за того, что траектория волнового пакета в суперпространстве может оказаться обращенной в динамическом «времени».

Пространственно-симметричные модели квантовой ОТО, которые мы обсуждали, имеют два общих недостатка. Их не так просто обобщить, чтобы они включали дополнительные несимметричные моды гравитационного поля; кроме того, эти модели имеют дело только с однородными вселенными. Вследствие последнего ограничения множество гамильтоновых связей общековариантной канонической теории (одна связь на каждую пространственную точку) передается одной связью. Таким образом, однородные модели не отражают многовременной аспект ОТО. Неоднородная модель, содержащая цилиндрически симметричные гравитационные волны постоянной поляризации, в настоящее время исследуется Кухаром. Эта модель особенно многообещающа вследствие своей бесконечномерности.

Для полного исправления этих дефектов требуется то, что мы называем неограниченной усеченной схемой. Основная черта такой схемы состоит в том, что она выражает полную ОТО через счетное число переменных вместо метрических коэффициентов $g_{ij}(P)$ в каждой пространственной точке. Такая схема должна дать правила для исключения всех переменных, кроме конечного набора динамических переменных. Полная теория тогда будет представляться *последовательностью* конечномерных усеченных теорий. Для расчета точного значения какой-либо численной величины вычисляем ее в каждой усеченной теории, а затем берем предел последовательности. В открытых вселенных S-матричный подход, описанный в § 6, является по крайней мере одним из кандидатов в усеченные схемы такого типа. Для этого в приведенном выше описании надо просто заменить слово «динамическая переменная» на «гравитон». Для замкнутых вселенных в настоящее время нет кандидата для неограниченной усеченной схемы, так как в полной нелинейной теории очень трудно разрешить пространственноподобные связи и разделить координатно зависимые и координатно независимые переменные. Необходимость использования полной нелинейной теории связана с тем, что в замкнутых вселенных нет асимптотических областей или радиационных зон, где можно применить линеаризацию.

Если неограниченная усеченная схема для ОТО будет получена, то возникнут некоторые упоминавшиеся выше трудности квантовой теории поля. Для данных коммутационных соотношений и данной усеченной схемы существует единственная квантовая теория. Более точно, существует последовательность единственных квантовых теорий. Но как выбрать усеченную схему? Имеется много способов выразить полевую теорию через счетную последовательность динамических переменных, и существует много возможных схем для исключения на данной стадии усечения нежелательных переменных. Даже классическая теория может дать следствия, зависящие от деталей усеченной схемы. Например, волновые числа звуковых волн в решетке с конечным периодом фиксируются только по модулю периодического упорядочения ячеек в пространстве волновых чисел. Волновые числа могут перескакивать от одной ячейки к другой, поэтому звуковые волны могут рассеиваться таким способом, который запрещен в непрерывной среде (рассеяние с перебросом). С уменьшением периода решетки растут волновые числа, при которых происходит аномальное рассеяние данного типа. Обычно берут усеченную схему, которая запрещает эти аномальные процессы, вводя импульсное обрезание или налагая условия непрерывности и дифференцируемости на поле, заменяющее решетку. Однако после того, как теория проквантована, очень трудно выяснить, какова естественная усеченная схема. В общем случае различные усеченные схемы дают унитарно неэквивалентные представления коммутационных соотношений. В настоящее время нет никаких оснований ожидать, что квантовая ОТО будет свободна от трудностей этого рода.

§ 8. СУПЕРПРОСТРАНСТВО

В этом параграфе мы возвращаемся к каноническому формализму квантовой гравитации и сосредоточим внимание на области определения функционалов квантовых состояний, т. е. на суперпространстве. Хотя это понятие присуще только одному представлению в одной формулировке теории, оказывается, что оно дает возможность подойти к некоторым наиболее образным и замечательным приложениям квантовой ОТО.

1. Что такое суперпространство?

Существует много способов классификации или представления функционала квантового состояния. Для каждого представления можно построить суперпространство, но обычно ограничиваются представлением метрики, как рассмотрено в § 5, п. 1 и в § 7, п. 2. Для данного трехмерного многообразия M можно построить пространство $\mathrm{Riem}(M)$ всевозможных римановых метрик, которые можно приписать многообразию. Однако точки этого пространства не находятся во взаимно однозначном соответствии с физическими состояниями, так как метрика содержит информацию не только о геометрии, но и о координатах, используемых для описания многообразия. Отождествим все точки (т. е. метрики) в $\mathrm{Riem}(M)$, связанные координатными преобразованиями, и получим предварительный вариант собственно суперпространства $S(M)$. Точное понятие суперпространства, наиболее полезное для квантовой геометродинамики, до настоящего времени все еще не определено. Ниже мы рассмотрим конечномерную модель суперпространства и укажем типичные его свойства, как они понимаются теперь, а также рассмотрим возможные модификации предварительного варианта суперпространства.

Вместо бесконечного многообразия метрик мы рассмотрим трехпараметрическое множество

$$ds^2 = f(\theta)\, ds_b^2 = a_1\left(1 + a_2 \sin\theta + a_3 \cos\theta\right) ds_b^2. \tag{8.1}$$

Здесь $ds_b^2 = d\theta^2 + d\phi^2 + d\chi^2$ — метрика на плоском трехмерном торе T^3; θ, ϕ, χ изменяются от 0 до 2π и для регулярности требуем $a_1 > 0$, $a_2^2 + a_3^2 < 1$. Мы называем это множество метрик $\mathrm{MRiem}(T^3)$ и рассматриваем его как простую модель пространства $\mathrm{Riem}(T^3)$ всех римановых метрик на T^3.

Пространство $\mathrm{MRiem}(T^3)$ не является моделью суперпространства $\mathrm{S}(T^3)$, так как каждая *геометрия* представлена в $\mathrm{MRiem}(T^3)$ много раз. Координатное преобразование

$$\theta' = \theta + \theta_0 \tag{8.2}$$

сохраняет форму (8.1) метрики, но изменяет параметры на

$$\begin{aligned} a_2' &= a_2 \cos\theta_0 - a_3 \sin\theta_0, \\ a_3' &= a_3 \cos\theta_0 + a_2 \sin\theta_0. \end{aligned} \tag{8.3}$$

Итак, все точки $\left(a_1, a_2', a_3'\right)$ в $\mathrm{MR}(T^3)$, получающиеся с помощью (8.3) из данной (a_1, a_2, a_3), представляют ту же самую геометрию. Такие точки называются орбитой координатного преобразования, проходящей через данную точку. В нашей модели орбитами являются окружности вокруг оси a_1 и точки на оси a_1; в общем случае орбиты будут замкнутыми подмногообразиями с размерностью, равной размерности группы координатных преобразований.

Для построения модели $\mathrm{MS}(T^3)$ суперпространства $\mathrm{S}(T^3)$ геометрий на T^3 мы образуем пространство всех орбит, т. е. отождествляем все метрики, отличающиеся только координатным преобразованием, и получаем пространство, где различаются только *геометрии*, а не метрики. Каждая орбита в $\mathrm{MS}(T^3)$ пересекается один раз с любой полоской единичной ширины, один край которой является осью a_1 (например, множество $\{a_i \mid 0 < a_2 < 1,\ a_3 = 0\}$). Следовательно, любое подпространство $\mathrm{MR}(T^3)$ есть представление $\mathrm{MS}(T^3)$. В качестве координат в $\mathrm{MS}(T^3)$ мы будем использовать a_1 и $r = \left(a_2^2 + a_3^2\right)^{1/2}$.

$\mathrm{MS}(T^3)$ содержит два типа точек: открытое множество $r > 0$ и «край» $r = 0$. Каждое из этих множеств является многообразием, но вместе они не образуют одного многообразия, так как в $\mathrm{MS}(T^3)$ точки на краю $r = 0$ имеют только «половинную» окрестность по сравнению с остальными точками. Следовательно, $\mathrm{MS}(T^3)$ не является многообразием, а состоит из двух множеств, или «слоев», каждое из которых есть многообразие. Это иллюстрирует общую ситуацию: любое суперпространство само не является многообразием, но расслаивается на последовательность многообразий [40]. Точки на одном слое соответствуют геометриям, имеющим бо́льшую симметрию, чем геометрии на следующем, большем слое. Как показывает наша модель, более симметричная геометрия имеет меньше соседних геометрий, чем менее симметричная [1]; следовательно, структура окрестностей последовательных слоев различается, и они не образуют одно многообразие.

Если мы считаем, что какой-то смысл имеет только 3-геометрия, а не любая дополнительная координатная информация, то мы должны принять суперпространство с его слоями. Однако де Витт [28] предположил, что слои необходимо убрать, а суперпространство превратить в многообразие, склеивая копии суперпространства вдоль слоев. В нашей модели мы должны склеить две полоски вдоль края $r = 0$, чтобы получить полоску удвоенной ширины, например подмножество $\mathrm{MRiem}(T^3)$, задаваемое

$$\{a_i \mid -1 < a_2 < +1,\ a_3 = 0\}.$$

Теперь каждая точка, включая точки на оси a_1, имеет полную окрестность, но все геометрии, за исключением симметричных геометрий на оси a_1, представлены дважды. На примере «миксмастеровского» мира мы уже видели расширенное суперпространство, где каждая миксмастеровская геометрия представлена шесть раз (§ 7, п. 4).

Расширенное суперпространство является одной из модификаций рассматриваемого понятия суперпространства. В дополнение к математическим преимуществам работы с одним многообразием имеются также физические мотивировки для такого расширения: кривые в суперпространстве, представляющие динамическое развитие пространства-времени (разрезанного на 3-геометрии при помощи некоторого временного координатного условия), являются гладкими в расширенном суперпространстве,

[1]) Сделаем вмятину на сфере; в результате независимо от того, где расположена эта вмятина, получится та же самая геометрия. Для яйца будет различие в зависимости от того, ближе к вершине или к основанию расположена вмятина. Геометрия яйца имеет больше соседних геометрий, чем более симметричная геометрия сферы.

тогда как в общем случае они могут отражаться под острыми углами от слоев самого суперпространства. Кроме того, малые негеометрические возмущения (типа слабых электромагнитных полей) могут быть однозначно описаны только в расширенном суперпространстве.

Другое расширение суперпространства предложил Уилер [93]. Он обратил внимание на тот факт, что классическую теорию можно сформулировать на многообразиях, обладающих различными топологиями. Можно показать [42], что классическое временное развитие запрещает топологии пространственноподобных поверхностей изменяться со временем; однако это не является достаточным основанием для утверждения, что такие изменения также запрещены квантовой динамикой. Вместо этого следует допустить возможность таких изменений и выяснить, связывает ли динамический закон, т. е. уравнение де Витта, состояния с различными топологиями с конечной амплитудой вероятности. Итак, наше понятие суперпространства будет оставаться неполным, пока мы не изучим возможность объединения геометрий с различными топологиями в единое суперпространство. Для трехмерного случая эта проблема сегодня еще ждет своего решения [1]). Однако Уилер дал основания надеяться, что в истинно квантовом состоянии будут происходить топологические изменения (см. п. 4 данного параграфа).

2. *Что дает суперпространство?*

Основная причина изучения суперпространства является также его наиболее важным достоинством: любой волновой функционал, определенный на суперпространстве, автоматически удовлетворяет пространственноподобным связям (5.18), поскольку эти связи требуют независимости волнового функционала от координат. Мы можем убедиться, что в нашей модели $MS(T^3)$ связи, генерирующие координатное преобразование (8.2), действительно удовлетворяются:

$$R_\theta = \int \frac{\partial x'^i}{\partial \theta_0} R_i d^3x = -\int \left(\frac{\partial x'^i}{\partial \theta_0}\right)_{;\,j} \pi^j_i d^3x = -\int \frac{df}{f\,\partial\theta} g_{ij} \frac{\delta}{\delta g_{ij}} d^3x. \quad (8.4)$$

Если $\psi[g_{ij}] = \psi(a_i)$, то благодаря уравнению (8.1) мы имеем

$$\frac{\partial \psi}{\partial a_k} = \int \frac{\partial g_{ij}}{\partial a_k} \frac{\delta \psi}{\delta g_{ij}} d^3x = \int \frac{\partial f}{f\,\partial a_k} g_{ij} \frac{\delta \psi}{\delta g_{ij}} d^3x. \quad (8.5)$$

---

[1]) Некоторые результаты для двумерного случая см. в работе Берса [12].

Если мы определим угол α с помощью соотношений $a_2 = r\cos\alpha$, $a_3 = r\sin\alpha$, то (8.1) показывает, что

$$\left.\frac{\partial f}{\partial \alpha}\right|_{a_1,\, r} = -\left.\frac{\partial f}{\partial \theta}\right|_{\phi,\, \chi},$$

поэтому для волнового функционала, определенного на $\mathrm{MRiem}(T^3)$, имеем

$$R_\theta = \frac{\partial}{\partial \alpha}. \tag{8.6}$$

Любой волновой функционал, зависящий только от a_1 и $r = (a_2^2 + a_3^2)^{1/2}$, удовлетворяет связи $R_\theta = 0$, но такие функции являются в точности функциями на модельном суперпространстве $\mathrm{MS}(T^3)$, так что любая функция ψ, определенная на $\mathrm{MS}(T^3)$, автоматически удовлетворяет $R_\theta = 0$. Поскольку эти связи в общем случае являются локальными уравнениями в $\mathrm{Riem}(M)$, они также удовлетворяются в любом расширенном суперпространстве.

Оставшаяся связь квантовой ОТО — гамильтонова связь, или уравнение де Витта $\mathscr{H}(x^i) = 0$ — полностью определяет динамику системы в суперпространстве, по крайней мере для замкнутых многообразий, где полный гамильтониан обращается в нуль. Уравнение де Витта (5.20) записано как одно уравнение в $\mathrm{Riem}(M)$ в пространственной точке (x^i). Если волновой функционал удовлетворяет суперпространственному условию симметрии (координатной инвариантности), то необходимо определить уравнение де Витта только в *одной* пространственной точке, так как пространственноподобные связи распространят его на все другие точки. Однако одно несимметричное уравнение для симметричной функции эквивалентно более чем одному симметричному уравнению. Таким образом, обширность суперпространства, содержащего геометрии для каждой мыслимой поверхности, компенсируется довольно ограниченным набором уравнений, определяющих динамику в суперпространстве.

Чтобы проиллюстрировать уравнение типа де Витта в модели $\mathrm{MRiem}(T^3)$, рассмотрим

$$H(\theta) = \sum_i \frac{\partial^2}{\partial a_i^2} + \cos\theta \frac{\partial^2}{\partial a_1\, \partial a_2} + \sin\theta \frac{\partial^2}{\partial a_1\, \partial a_3}. \tag{8.7}$$

Если $H(\theta)\psi$ обращается в нуль при $\psi = \psi(a_1, r)$ для одного значения θ, то оно обращается в нуль при всех значениях θ, но этот факт заключает в себе два уравнения для ψ:

$$\left(\frac{\partial^2}{\partial a_1^2} + \frac{\partial^2}{\partial r^2} + \frac{\partial}{r\,\partial r}\right)\psi = 0, \quad \frac{\partial^2 \psi}{\partial a_1\, \partial r} = 0.$$

При этом $H(\theta)$ не рассматривают как приближение к оператору де Витта, и такой приближенный оператор в $\mathrm{MRiem}(T^3)$ не очень полезен. Это модельное пространство слишком мало, чтобы дать реальную картину динамики в суперпространстве.

Концепции суперпространства особенно полезны для понимания связи между классической физикой 4-геометрий и, по-видимому, совершенно отличной квантовой физикой 3-геометрий. Рассмотрим классическое 4-пространство, удовлетворяющее уравнениям Эйнштейна. Любая пространственноподобная поверхность, лежащая в этом пространстве, несет 3-геометрию и, следовательно, соответствует точке в суперпространстве. Множество («классический каталог») всех таких точек, получаемых при взятии всевозможных пространственноподобных сечений 4-геометрии, является суперпространственным представлением этой 4-геометрии. Это аналог пути классической частицы в пространстве-времени. Но если путь частицы определяется положением *и* временем, то в суперпространственном описании 4-геометрии временная координата не нужна: информации о 3-геометрии на поверхности достаточно, чтобы определить, где эта поверхность должна располагаться в массе 3-геометрий, т. е. в пространстве-времени [7].

Типичное квантовое состояние распространяется на области суперпространства, гораздо большие, чем классический каталог. Чтобы получить классический предел, необходимо построить узкие волновые пакеты, как в обычной квантовой механике. Такие волновые пакеты должны иметь пик в классическом каталоге и становиться пренебрежимо малыми вне его. В классическом пределе можно в качестве компонент волн взять решения наинизшего ВКБ-приближения уравнения де Витта. Фаза этих волн удовлетворяет уравнению Гамильтона — Якоби ОТО (см. § 7, п. 3), и было показано [41], что строго на классическом каталоге существует интерференция, так что классическим пределом квантовой динамики действительно является классическая динамика, определяемая уравнениями Эйнштейна.

3. Метрики для суперпространства

Мы видели, что (расширенное) суперпространство имеет структуру многообразия. Таким образом, обычные понятия функций (например, функционала состояния), дифференциальных операторов (например, дифференциального оператора гамильтоновой связи), касательных пространств и т. д. имеют смысл в суперпространстве. Любое многообразие можно наделить римановой метрикой, и совершенно естественно рассмотреть те суперпространственные метрики, которые имеют какой-то физический смысл.

Объекты суперпространства в общем случае получаются из координатно инвариантных объектов в Riem(M). Например, касательное пространство Riem(M) определяется голономным базисом $\delta/\delta g_{ij}(x)$, поскольку в каждой пространственной точке x в M существуют шесть касательных векторов. Суперпространство имеет только три касательных вектора в пространственной точке, так как три касательных вектора в пространственной точке $(\delta/\delta g_{ij})_{;j}$ соответствуют координатным преобразованиям и, следовательно, отождествляются с нулевым вектором в касательном к суперпространству пространстве. В нашей модели MRiem(T^3) касательные пространства трехмерны и определяются $\partial/\partial a_i$, но касательные пространства MS(T^3) двумерны и определяются $\partial/\partial a_1$, $\partial/\partial r$; касательная к орбитам $-a_3(\partial/\partial a_2)+ a_2(\partial/\partial a_3)=\partial/\partial\alpha$ есть пространственноподобная связь (8.6), которая автоматически равна нулю в MS(T^3).

Аналогично можно получить имеющую физический смысл суперпространственную метрику, определив сначала метрику в Riem(M), а затем выделяя ее координатно инвариантные свойства. Такая метрика G, как и в других квантовых теориях, определяется членом, являющимся «кинетической энергией» суммарного гамильтониана $H=\int\mathscr{H}(x)\,d^3x$. Этот член имеет форму

$$\int\int G_{ijkl}(x,x')\frac{\delta^2}{\delta g_{ij}(x)\,\delta g_{kl}(x')}\,d^3x d^3x',$$

где ковариантные метрические коэффициенты равны

$$G_{ijkl}(x,x')=\frac{1}{2}\delta(x,x')\,g^{-1/2}(g_{ik}g_{jl}+g_{il}g_{jk}-g_{ij}g_{kl}). \quad (8.8)$$

Соответствующая контравариантная метрика есть обратная величина

$$\left\langle\frac{\delta}{\delta g_{ij}},\frac{\delta}{\delta g_{kl}}\right\rangle=G^{ijkl}=\frac{1}{2}\delta(x,x')\,g^{1/2}(g^{ik}g^{jl}+g^{il}g^{jk}-2g^{ij}g^{kl}). \quad (8.9)$$

Де Витт проанализировал структуру этой метрики. Она диагональна относительно метрических вариаций в различных пространственных точках x, x'; в данной пространственной точке она определяет гиперболическое скалярное произведение метрических вариаций с одним «времениподобным» и пятью «пространственноподобными» направлениями. «Времениподобная» вариация соответствует общему расширению или сжатию, в то время как вариации в «пространственноподобных» направлениях оставляют элемент объема $(\det g)^{1/2}$ постоянным. Сама метрика $\langle\ldots\rangle$

является координатным инвариантом, так как координатная зависимость в обеих частях (8.9) одинакова. Экстремальное, или нормальное, расстояние между смежными орбитами в $\mathrm{Riem}(M)$ можно взять в качестве расстояния в $S(M)$.

Метрика (8.9) становится сингулярной, когда фактор $g^{1/2}$ обращается в нуль. Такие сингулярные точки в $\mathrm{Riem}(M)$ находятся на конечном расстоянии от регулярных точек и образуют «границу» этого многообразия. Суперпространство $S(M)$ имеет соответствующую границу для сингулярных геометрий. В нашей модели $MS(T^3)$ граница состоит из линий $a_1 = 0$ и $|r| = 1$. Теоремы о наличии сингулярностей (коллапса) в общем классическом решении (см. § 2, п. 1) означают на языке суперпространства, что классические множества C в общем случае имеют предельные точки на границе.

Можно ожидать, что в граничной области, где кривизна становится большой и преобладают квантовые эффекты, квантовое поведение совершенно отличается от классического. Единственный путь гарантировать отсутствие сингулярностей в квантовой ситуации состоит в наложении граничного условия — обращения в нуль функционала состояния Ψ на границе сингулярности, как предложил де Витт для квантованной вселенной Фридмана [1]). Граничное условие и гамильтонова связь могут слишком ограничить или даже исключить все, кроме очень немногих, возможно кроме единственного, функционалы состояний. Некоторых авторов [24, 32, 91] привлекает идея единственности функции состояния, так как тогда единственность реальной вселенной отражалась бы в математической теории. Однако другие авторы [67] предпочитают менее жесткие граничные условия и рассматривают квантовую динамику больше как проблему рассеяния, чем проблему связанного состояния (см. также § 7, п. 4).

В дополнение к «кинетической» метрике предлагались другие метрики, которые учитывают член потенциальной энергии в уравнении де Витта. Выбирая определенные координатные условия, можно определить, как расщепить любое пространство-время на частную последовательность 3-геометрий. Любое классическое решение теперь соответствует (одномерной) кривой в $S(M)$. Были найдены метрики [28, 45], в которых эти классические кривые являются геодезическими. Такие метрики, в частности, полезны при изучении динамики суперпространства.

---

[1]) Поскольку сингулярные 3-геометрии могут возникать на пространственноподобных поверхностях вполне регулярных 4-геометрий, более подходящее обобщение граничного условия де Витта должно содержать связь между внутренней геометрией и внешней кривизной.

Например, динамические сингулярности появляются в этих метриках для «симметричных по времени» геометрий при исчезновении ^3R, которые кинематически не являются сингулярными. Таким образом, такие суперпространственные метрики могут помочь обнаружить дальнейшие физически важные особенности суперпространства и функционала состояния.

4. Какая физика заключена в суперпространстве?

Мы уже упоминали об одном важном моменте квантовой динамики в суперпространстве — избежании граничных сингулярностей. Это квантовое явление должно проявляться на далекой стадии коллапса. Более локальные квантовые эффекты имеют место на всех стадиях эволюции вселенной и являются наиболее значительными в областях с размерами, сравнимыми с планковской длиной. Уилер [90, 93] описал качественно некоторые из явлений, появление которых ожидается при таком квантовом режиме.

Планковская длина $(\hbar G/c^3)^{1/2} \sim 1{,}6 \cdot 10^{-33}$ см определяет масштаб квантовой гравитации [73]. Это размер, до которого может происходить гравитационный коллапс, прежде чем его смогут остановить квантовые эффекты. Это также размер, ниже которого становятся большими квантовые «флуктуации» геометрии. Большие квантовые неопределенности в геометрии $\delta g_{ij} \sim 1$ Уилер интерпретирует как указание на то, что даже топологию пространства-времени нельзя хорошо определить в таких малых областях и что ее можно представить себе резонирующей между многими различными возможностями. В малых масштабах пространство-время напоминает «пену», в свойствах и динамике которой важную роль может играть топологическая тонкая структура.

Размеры обычных элементарных частиц огромны по сравнению с тонкостью пены. Поскольку мы не знаем механизма, способного остановить коллапс, прежде чем он достигнет таких малых размеров, можно думать, что элементарные частицы не переживут коллапса и только сама пена останется целой. Эта точка зрения подтверждается классическими расчетами, которые показывают, что большинство свойств элементарной частицы исчезнет, когда она попадет в шварцшильдовскую «черную дыру». Если частицы и поля могут исчезнуть в пене, то разумно предположить, что на «последующей» стадии расширения вселенной они могут возникнуть из пены. Таким образом, с точки зрения Уилера, массы, заряды и другие свойства элементарных частиц не предопределены в динамике вселенной, а являются следствием того, что произошло очень рано при образовании вселенной (при «большом взрыве»).

Тогда возникают следующие важные вопросы относительно суперпространственной динамики пены: можно ли построить устойчивую структуру из флуктуирующей геометрии; существует ли некоторый механизм, делающий эти структуры достаточно однородными, чтобы они могли представлять собой элементарные частицы; можно ли рассматривать электромагнитные и другие поля как аналогичные крупномасштабные аспекты квантовой геометрии, которые теряют свой смысл на расстояниях порядка планковской длины; можно ли найти какую-то причину того, что в настоящем цикле эволюции вселенной мы находим в ней данные конкретные элементарные частицы? Вопросы, аналогичные этим, могут руководить нами в поиске новых физических явлений, связанных с суперпространством.

Одним из самых центральных моментов в динамике суперпространства является интерпретация функционала состояния. Если он на самом деле единствен, то это фактически приводит к интерпретации Эверетта [32] (см. также [27]): одно состояние Ψ содержит описание как вселенной, так и наблюдателей. «Относительное состояние» каждого наблюдателя является компонентой Ψ, содержащей информацию, доступную этому наблюдателю. Полное состояние Ψ содержит суперпозицию всевозможных различных вселенных на каждой стадии их эволюции, последовательно скоррелированных с их соответствующими наблюдателями и связанных друг с другом в масштабе планковской длины. Таким образом, одно Ψ уже содержит все следствия квантовой динамики. Вместе с Уилером [94] можно надеяться, что в конце концов мы придем к объекту, подобному этому Ψ, который содержит всю динамику, но сам не выделяется динамическим уравнением, а следует из некоторого простого, необходимого и красивого принципа.

§ 9. ЗАКЛЮЧЕНИЕ

В начале этой статьи мы изобразили физику как совокупность разрозненных областей, в которой часть, обозначенная ОТО, располагалась отдельно от остальных. Подводя итог описанным в этой статье результатам, можно сказать, что различные области удалось частично сблизить. Часть, обозначенная ОТО, теперь связана с остальными. В § 3 и 4 показано, как ОТО переводится в рамки канонического формализма, который выявляет ее существенное сходство с другими классическими теориями поля. В § 5 мы видели, что к ОТО можно применить ряд методов квантования. Теперь уже не существует какой-либо трудности в *построении* квантовой ОТО. К сожалению, имеется ряд различных путей построения такой теории и каждый из этих подходов содержит элементы неоднозначности (упорядочение сомножителей, выбор канонических переменных, выбор меры, физическая интерпретация, исключение бесконечностей и т. д.). Конечно, некоторые из этих неоднозначностей имеются и в негравитационных квантовых теориях и либо были разрешены с помощью эксперимента, либо не решались вообще, как в случае неперенормируемых теорий поля. Разрозненные части физической картины теперь можно собрать вместе, но они все же не образуют единого целого.

Следующие этапы в исследовании квантовой ОТО должны прояснить физический смысл теории, например путем проведения фактических расчетов для физически интересных квантовых процессов. В конце концов гравитационные и негравитационные квантовые теории должны возникнуть как различные аспекты единой квантовой теории. Будущее решит, будет ли эта теория больше похожей на современную квантовую гравитацию с другими полями, представленными как особенности квантовой геометрии, или она будет ближе к современным квантовым теориям поля с гравитацией, играющей более тонкую роль (например, роль универсального регулятора [23]). Благодаря своей необычной природе и чрезвычайно слабому взаимодействию классическая теория гравитации в течение многих лет стояла в стороне от основного русла развития физики. Надежда на то, что именно эти особенности покажут, что квантовая гравитация представляет собой существенную составную часть последовательной квантовой теории, является одним из сильнейших стимулов дальнейших исследований в области квантовой общей теории относительности.

ЛИТЕРАТУРА

1. *Adler R., Bazin M., Schiffer M.*, Introduction to General Relativity, New York — London, 1965.
2. *Anderson J. L.*, Principles of Relativity Physics, New York, 1967.
3. *Anderson J. L., Bergmann P. G.*, Phys. Rev., **83**, 1018 (1951).
4. *Araki H.*, Ann. Phys., N. Y., **7**, 456 (1959).
5. *Arnowitt R., Deser S., Misner C. W.*, Gravitation: An Introduction to Current Research, ed. L. Witten, New York — London, 1962.
6. *Aslaksen E. W.*, Ph. D. Thesis, Lehiga University, Bethlehm, Pa., 1968.
7. *Baierlein R. F., Sharp D. H., Wheeler J. A.*, Phys. Rev., **126**, 1864 (1962).
8. *Barker B. M., Bhatia M. S., Gupta S. N.*, Phys. Rev., **182**, 1387 (1969).
9. *Barker B. M., Gupta S. N., Kaskas J.*, Phys. Rev., **182**, 1391 (1969).
10. *Bergmann P. G.*, Phys. Rev., **75**, 680 (1949).

11. *Bergmann P. G.*, Nuovo Cimento, **3**, 1177 (1956).
12. *Bers L.*, Analytic Methods in Mathematical Physics, ed. R. P. Gilbert and G. Newton, New York, 1970.
13. *Bishop R. L.*, *Crittenden R. J.*, Geometry of Manifolds, New York — London, 1964 (см. перевод: *Р. Л. Бишоп*, *Р. Дж. Криттенден*, Геометрия многообразий, изд-во «Мир», М., 1967).
14. *Bjorken J. D.*, Phys. Rev., **148**, 1467 (1966).
15. *Боголюбов Н. Н.*, *Ширков Д. В.*, Введение в теорию квантованных полей, М., 1957.
16. *Bohr N.*, *Rosenfeld L.*, Mat.-Fys. Medd., **12**, **№ 8** (1933) (см. перевод: *Н. Бор*, Собрание сочинений, том II, М., 1971).
17. *Brill D. R.*, Ann. Phys., N. Y., **7**, 466 (1959).
18. *Brill D. R.*, *Deser S.*, Ann. Phys., N. Y., **50**, 548 (1968).
19. *Brill D. R.*, *Hartle J. B.*, Phys. Rev., **135**, B271 (1964).
20. *Deser S.*, *Laurent B. E.*, Ann. Phys., N. Y., **50**, 76 (1968).
21. *DeWitt B. S.*, Journ. Math. Phys., **3**, 1073 (1962).
22. *DeWitt B. S.*, в книге Relativity, Groups and Topology, ed. B. S. DeWitt and C. M. DeWitt, New York — London, 1964.
23. *DeWitt B. S.*, Phys. Rev. Lett., **13**, 114 (1964).
24. *DeWitt B. S.*, Phys. Rev., **160**, 1113 (1967).
25. *DeWitt B. S.*, Phys. Rev., **162**, 1195 (1967).
26. *DeWitt B. S.*, Phys. Rev., **162**, 1239 (1967).
27. *DeWitt B. S.*, в книге Battelle Recontres: 1967 Lectures in Mathematics and Physics, ed. C. M. DeWitt and J. A. Wheeler, New York — Amsterdam, 1968, p. 318.
28. *DeWitt B. S.*, Relativity: Proceedings of the Relativity Conference in the Midwest, ed. L. Witten, New York, 1970.
29. *Dirac P. A. M.*, Can. Journ. Math., **2**, 129 (1950).
30. *Dirac P. A. M.*, Proc. Roy. Soc., **A246**, 333 (1958).
31. *Dirac P. A. M.*, Phys. Rev., **114**, 924 (1959).
32. *Everett H.*, Rev. Mod. Phys., **29**, 454 (1957).
33. *Fadeev L. D.*, *Popov V. N.*, Phys. Lett., **25**, B29 (1967).
34. *Fermi E.*, Rev. Mod. Phys., **4**, 87 (1932) (см. перевод: *Э. Ферми*, Собрание сочинений, том 1, М., 1971).
35. *Feynman R. P.*, Ph. D. Thesis, Princeton University, Princeton, N. J., USA, 1942.
36. *Feynman R. P.*, Phys. Rev., **76**, 769 (1949) (см. перевод в сборнике «Новейшее развитие квантовой электродинамики», ИЛ, М., 1954).
37. *Feynman R. P.*, в книге Report of the Conference on the Role of Gravitation in Physics, Chapel Hill, North Carolina WAPC Tech. Rept. 57—216, U. S. Dept. of Commerce, Washington, D. C., 1957.
38. *Feynman R. P.*, Acta Phys. Polon, **24**, 697 (1963).
39. *Feynman R. P.*, *Hibbs A. R.*, Quantum Mechanics and Path Integrals, New York — London, 1965 (см. перевод: *Р. Фейнман*, *А. Хибс*, Квантовая механика и интегралы по траекториям, изд-во «Мир», М., 1968).
40. *Fischer A. E.*, Ph. D. Thesis, Princeton University, Princeton, N. J., USA, 1969.
41. *Gerlach U. H.*, Phys. Rev., **177**, 1929 (1969).
42. *Geroch R. P.*, Journ. Math. Phys., **8**, 782 (1967).
43. *Glashow S. L.*, *Gell-Mann M.*, Ann. Phys., N. Y., **15**, 437 (1961) (см. перевод в сборнике «Элементарные частицы и компенсирующие поля», изд-во «Мир», М., 1964).
44. *Godfrey S.*, Ph. D. Thesis, Yale University, New Haven, Conn., USA, 1968.
45. *Gowdy R. H.*, Phys. Rev. (1970).
46. *Guggenheimer H. W.*, Differential Geometry, New York — London, 1963.

47. *Gupta S. N.*, Phys. Rev., **172**, 1303 (1968).
48. *Hawking S. W.*, Proc. Roy. Soc., **A300**, 187 (1967).
49. *Hicks N. J.*, Notes on Differential Geometry, Princeton, N. J., 1965.
50. *Higgs P. W.*, Phys. Rev. Lett., **3**, 66(E) (1959).
51. *Hori S.*, Prog. Theor. Phys., **7**, 589 (1952).
52. *Johnson K.*, Nucl. Phys., **25**, 431 (1961).
53. *Klauder J. R.*, Ann. Phys., N. Y., **11**, 123 (1960).
54. *Klauder J. R.*, Nuovo Cimento, **19**, 1059 (1961).
55. *Klauder J. R.*, Nuovo Cimento, **25**, 542 (1962).
56. *Klauder J. R.*, Journ. Math. Phys., **9**, 206 (1968).
57. *Klauder J. R.*, Relativity: Proc. Midwest Conf. Relativity, ed. L. Witten, New York, 1969.
58. *Komar A. B.*, Phys. Rev., **111**, 1182 (1958).
59. *Komar A. B.*, Phys. Rev., **164**, 1595 (1967).
60. *Ландау Л. Д., Лифшиц Е. М.*, Теория поля, изд-во «Наука», М., 1967.
61. *Laurent B. E.*, Ark. Fys., **16**, 279 (1959).
62. *Lee T. D.*, Phys. Rev., **171**, 1731 (1968).
63. *Leutwyler H.*, Phys. Rev., **134**, 1155 (1964).
64. *Mandelstam S.*, Ann. Phys., N. Y., **19**, 1 (1962).
65. *Mandelstam S.*, Phys. Rev., **175**, 1580 (1968).
66. *Misner C. W.*, Rev. Mod. Phys., **29**, 497 (1957).
67. *Misner C. W.*, Phys. Rev., **186**, 1319 (1969).
68. *Misner C. W.*, Relativity: Proc. Midwest Conf. Relativity, ed. L. Witten, New York, 1970.
69. *Moller C.*, The Theory of Relativity, London, 1962.
70. *Novozilov J. V.*, *Tulub A. V.*, Fortschr. Phys., **6**, 50 (1958).
71. *Penrose R.*, в книге Relativity, Groups and Topology, ed. B. S. DeWitt and C. M. DeWitt, New York, 1964 (см. перевод в сборнике «Гравитация и топология», изд-во «Мир», М., 1966).
72. *Peres A.*, Nuovo Cimento, **26**, 53 (1962).
73. *Planck M.*, The Theory of Heat Radiation, Philadelphia, 1914 (reprinted, New York, 1959).
74. *Radkowski A.*, Ph. D. Thesis, Harvard University, Cambridge, Mass., USA, 1968.
75. *Roman P.*, Introduction to Quantum Field Theory, New York — London, 1969.
76. *Rosenfeld L.*, Zs. Phys., **65**, 589 (1930).
77. *Rosenfeld L.*, Nucl. Phys., **40**, 353 (1963).
78. *Russo R.*, Ph. D. Thesis, Northeastern University, Boston, Mass., USA, 1969.
79. *Sachs R. K.*, Phys. Rev., **128**, 2851 (1962).
80. *Sachs R. K.*, в книге Relativity, Groups and Topology ed. B. S. DeWitt and C. M. DeWitt, New York — London, 1964 (см. перевод в сборнике «Гравитация и топология», изд-во «Мир», М., 1966).
81. *Sakurai J. J.*, Advanced Quantum Mechanics, Reading, Mass., 1967.
82. *Schweber S. S.*, An Introduction to Relativistic Quantum Field Theory, New York, 1962 (см. перевод: *С. Швебер*, Введение в релятивистскую квантовую теорию поля, ИЛ, М., 1963).
83. *Schwinger J.*, Phys. Rev., **74**, 1449 (1948).
84. *Schwinger J.*, Phys. Rev., **152**, 1219 (1966).
85. *Schwinger J.*, Phys. Rev., **158**, 1391 (1967).
86. *Schwinger J.*, Phys. Rev., **173**, 1264 (1968).
87. *Thomas L.*, Ph. D. Thesis, Yale University, New York, New Haven, Conn., 1970.

88. *Tomonaga S.*, Progr. Theor. Phys., **1**, 27 (1946) (см. перевод в сборнике «Новейшее развитие квантовой электродинамики», ИЛ, М., 1954).
89. *Weinberg S.*, Phys. Rev., **140**, 516 (1965).
90. *Wheeler J. A.*, Ann. Phys., N. Y., **2**, 604 (1957).
91. *Wheeler J. A.*, Rev. Mod. Phys., **29**, 463 (1957).
92. *Wheeler J.*, в книге Relativity, Groups and Topology, ed. B. S. DeWitt and C. M. DeWitt, New York — London, 1964.
93. *Wheeler J. A.*, в книге Battelle Recontres: 1967 Lectures in Mathematics and Physics, ed. C. M. DeWitt and J. A. Wheeler, New York — Amsterdam, 1968.
94. *Wheeler J. A.*, в книге Relativity: Proceedings of the Relativity Conference in the Midwest, ed. L. Witten, New York, 1970.
95. *Weyl H.*, Phys. Rev., **77**, 699 (1950).
96. *Kibble T. W. B.*, Journ. Math. Phys., **2**, 212 (1961).
97. *Фролов Б. Н.*, в сборнике «Современные проблемы гравитации», Тбилиси, 1967, стр. 270.
98. *Фролов Б. Н.*, Тезисы докладов 3-й Советской гравитационной конференции, Ереван, 1972, стр. 170.

ПОСВЯЩАЕТСЯ ПАМЯТИ Л. ШИФФА

Статья 3

Вычисление перенормировочных констант

А. САЛАМ

Abdus Salam, обзорный доклад,
Coral Gables Conference on Fundamental Interactions at High Energy,
1971 ICTP preprint, Trieste IC/71/3.

ВВЕДЕНИЕ

Темой настоящего доклада является вычисление традиционно бесконечных перенормировочных констант. Эти константы обычно выражаются (в формализме Челлена — Лемана) в виде интегралов от спектральных функций и их моментов. Поэтому они представляют собой величины, измеримые или неизмеримые, в зависимости от соответствующих формфакторов. Для теорий, обладающих внутренней симметрией, в действительности мы знаем о них несколько больше. Например, мы полагаем, что голые (затравочные) значения m_{π^+} и m_{π^0} равны; аналогично равны голые значения g_A и g_V, а также (вероятно) m_e и m_ν $(=0)$.

При вычислении перенормировочных констант мы сталкиваемся с произведением сингулярных распределений (сингулярных обобщенных функций). Основная проблема заключается в выделении хорошей физики из плохой математики. В настоящее время это оказывается возможным благодаря следующим успехам в теории.

1. *Успехи в математике* обобщенных функций. Я имею в виду работы Гельфанда — Шилова и связанные с ними работы, посвященные определению фурье-образов произведений распределений типа

$$\left(\frac{1}{x^2}\right)^{z_1} \otimes \left(\frac{1}{x^2}\right)^{z_2} \quad \text{и} \quad \partial_\mu \left(\frac{1}{x_2}\right)^{z_1} \otimes \partial_\nu \left(\frac{1}{x^2}\right)^{z_2}$$

для целых z. Эти методы используют принцип *аналитического* продолжения (по переменным z_1 и z_2 из области $0 < \operatorname{Re} z_1$, z_2, $z_1 + z_2 < 1$) и представляют собой главный и *до сих пор еще не оцененный* успех в математическом аспекте рассматриваемой проблемы.

2. *Успехи в теории поля.* Имеются в виду следующие достижения:

а) обоснование утверждения, что все представляющие физический интерес лагранжианы *по существу неполиномиальны* (даже если в некоторых случаях это утверждение представляется неверным, включение гравитационного взаимодействия все равно делает все лагранжианы неполиномиальными);

б) доказательство (которое будет здесь представлено) того факта, что локализуемые неполиномиальные лагранжианы заслуживают такого же доверия в строгом теоретико-полевом смысле, как и полиномиальные;

в) проверка того факта, что упомянутые выше аналитические регуляризационные методы *абсолютно приспособлены* для подобных лагранжианов и позволяют получить для целого ряда перенормировочных констант *конечные* и *однозначные* значения (при некотором добавочном физическом допущении, как будет доложено проф. Леманом).

Наш обзорный доклад делится на три части:

Часть I. Аналитические регуляризационные методы и их применимость к неполиномиальным лагранжианам.

Часть II. Эффекты включения в электродинамику лептонов квантованной *тензорной* гравитации, т. е. *конечное и калибровочно инвариантное вычисление собственного заряда и собственной массы электрона* в квантовотеоретической трактовке искривленного пространства-времени. (Сохранение *калибровочной инвариантности* представляет собой новый результат.)

Часть III. Гипотетическое предположение о том, что связь F-мезонов с адронным тензором энергии имеет тот же (неполиномиальный) вид, что и связь эйнштейновских гравитонов с лептонами (т. е. введение постулата о двутензорной теории гравитации), и возможность использования указанной неполиномиальности для регуляризации перенормировочных констант в физике сильных взаимодействий.

В своих вступительных замечаниях я коротко остановлюсь на некоторых новейших результатах, предположениях и нерешенных вопросах. (В докладах, подобных этому, можно касаться самых невероятных предположений и домыслов с целью обратить внимание на возможные направления дальнейшего развития теории.) Однако, прежде чем перейти к основной части, полезно, по-видимому, привести некоторые неполиномиальные лагранжианы, играющие важную роль в физике.

Представляют физический интерес следующие неполиномиальные лагранжианы:

А. Киральный $SU(2) \times SU(2)$-лагранжиан сильных взаимодействий

Типичным примером служит π-мезонный лагранжиан в различных способах параметризации:

$$L = \mathrm{Sp}\, \partial_\mu S \partial_\mu S^+,$$

где

$$S = \frac{1 + i\lambda_W \boldsymbol{\tau} \cdot \boldsymbol{\pi}}{1 - i\lambda_W \boldsymbol{\tau} \cdot \boldsymbol{\pi}}$$

(в координатах Вейнберга — Швингера) или

$$S = \exp(i\lambda_G \boldsymbol{\tau} \cdot \boldsymbol{\pi})$$

(в координатах Гюрши). Здесь константа λ (λ_W или λ_G), носящая название *малой константы связи*, имеет размерность обратной длины (эмпирически $\lambda \approx m_\pi^{-1}$). Важный открытый вопрос в теории поля заключается в общем виде в следующем. Совпадают ли для этих двух вариантов элементы S-матрицы на массовой оболочке? Это особенно важно, поскольку (экспоненциальная) форма Гюрши, по-видимому, определяет локализуемую теорию π-мезонов, а (рациональная) форма Вейнберга — нелокализуемую (см. приложение).

Б. Лагранжиан слабых взаимодействий с промежуточным бозоном

Типичным примером служит нейтральный W-мезон массы m, взаимодействующий с кварками Q массы M; лагранжиан взаимодействия имеет вид

$$L_{\text{вз}} = f\bar{Q}\gamma_\mu (1 + \gamma_5) Q W_\mu,$$

где $m^2 f^2 \approx G_F$ (константа Ферми). Существенная неполиномиальность теории заключена в связи с производной поля спина 0, «дочернего» по отношению к физической частице спина 1, описываемой 4-векторным полем W_μ. Чтобы пояснить сказанное, запишем

$$W_\mu = A_\mu + \frac{1}{m} \partial_\mu B$$

в хорошо известной форме Штюкельберга и совершим преобразование кваркового поля

$$Q' = \exp\left(\frac{if\gamma_5 B}{m}\right) Q.$$

Преобразованный лагранжиан взаимодействия имеет вид

$$L_{\text{вз}} = f\bar{Q}'\gamma_\mu(1+\gamma_5)\,Q'A_\mu + M\bar{Q}'\left(\exp\left[\frac{2if\gamma_5 B}{m}\right]-1\right)Q'.$$

Константа $f/m \approx \sqrt{G_F}$ во втором члене этого лагранжиана играет роль малой константы связи. Важный момент, который я хотел бы подчеркнуть, заключается в следующем: *связь через производную* (в данном случае от B-поля) *может обманчиво выглядеть полиномиальной*; ее существенная неполиномиальность (с характерным свойством, заключающимся на языке Фейнмана в том, что из одной вершины исходит множество линий) может быть сделана явной подходящим преобразованием полевых переменных.

В. Эйнштейновская тензорная гравитация и взаимодействующие с гравитацией поля обычной материи

Общепринятый гравитационный лагранжиан имеет вид

$$L_{\text{эйншт}} = \varkappa^{-2} g^{\mu\nu}\left(\Gamma^\lambda_{\mu\rho}\Gamma^\rho_{\nu\lambda} - \Gamma^\lambda_{\mu\nu}\Gamma^\rho_{\lambda\rho}\right)/\sqrt{-\det(g^{\mu\nu})},$$

где

$$\Gamma^\lambda_{\mu\nu} = \frac{1}{2} g^{\lambda\rho}\left(\partial_\mu g_{\nu\rho} + \partial_\nu g_{\mu\rho} - \partial_\rho g_{\mu\nu}\right).$$

Если $g^{\mu\nu}$ — фундаментальное поле, то ковариантное поле $g_{\mu\nu}$ существенно неполиномиально, и наоборот.

Простейшим примером поля обычной материи в искривленном пространстве служит скалярное поле

$$L_{\text{материи}} = \frac{g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi}{\sqrt{-\det g^{\mu\nu}}}.$$

Этот лагранжиан также неполиномиален. Величина $g^{\mu\nu}$ (метрический тензор классической физики) обычно параметризуется следующим образом (когда пространство-время на бесконечности минковское):

$$g^{\mu\nu} = \eta^{\mu\nu} + \varkappa h^{\mu\nu}, \quad \eta^{\mu\nu} = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}.$$

Здесь $\varkappa^2$, константа связи теории, равна $8\pi G_N$ (G_N — ньютонова константа) $\approx 10^{-44}/m_e^2$, т. е. $\varkappa^{-1} \approx 10^{18}$ БэВ. Альтернативная (и предпочитаемая математиками) параметризация задается выражением

$$g^{\mu\nu} = \left[\exp\left(\varkappa\gamma_{ab}h^{ab}\right)\right]^{\mu\nu},$$

где γ_{ab} — псевдосимметричная матрица 4×4. Отметим, что для этой «экспоненциальной» параметризации

$$\det g^{\mu\nu} = \exp\left(\varkappa h_a^a\right).$$

Когда мы приступаем к изучению частиц спина $^1/_2$, тензор Эйнштейна нельзя рассматривать как фундаментальное поле. Вместо этого можно работать с тетрадной гравитацией, в которой основную роль играют величины $L^{\mu a}$, связанные с $g^{\mu\nu}$ соотношением $g^{\mu\nu} = L^{\mu a}L_a^{\nu}$.

ЧАСТЬ I

Эта часть представлена на конференции тремя докладами.

Проф. Тейлор в своем докладе сделает обзор методов, развитых для вычисления элементов S-матрицы в неполиномиальных лагранжевых теориях для любого желаемого порядка по главной константе связи и для всех порядков по малой константе связи. (Напомню, что в лагранжиане $L_{вз} = g\!:\!\exp(\varkappa\phi) - 1\!:$ константу g мы называем главной, а константу $\varkappa$ — малой.) Проф. Тейлор проиллюстрирует построение механизма подавления бесконечностей для рассматриваемых теорий и даже приведет прекрасное новое доказательство унитарности этого решения. Проф. Леман в своем докладе обсудит очень важную проблему *обобщенно-функциональной неоднозначности* и ее устранения в определении упорядоченных по времени произведений в *локализуемых теориях*. Эти результаты вытекают из работы Лемана и Польмейера, которые показали, что процедура, развитая Филипповым, Волковым, Саламом, Стрэтди и др., вполне гарантирует аналитичность и унитарность локализуемых теорий в любом порядке по главной константе. Демонстрация того, что для локализуемых неполиномиальных теорий аналитичность и унитарность выполняются столь же хорошо, как и для общепринятых полиномиальных теорий, оказалась для меня потрясающей новостью. Ефимов уже дал одно доказательство этого факта; более строгое подтверждение Лемана и Польмейера (и для унитарности Тейлора) можно всячески приветствовать. В математическом отношении во всяком случае рассматриваемые теории настолько «респектабельны» (т. е. обычны и нормальны, лишены какой-либо окутывающей их тайны), насколько можно было бы желать. За докладом проф. Лемана следует доклад д-ра Криста, обсуждающего проблему однозначности с несколько отличной точки зрения.

Я дам здесь очень краткий обзор идей, которые более детально рассматриваются в этих докладах.

А. Метод Гельфанда — Шилова и подавление бесконечностей

1. *Проблемы.* Задавая локализуемый лагранжиан вида

$$L_{\mathrm{I}} = g\,[\exp(\varkappa\phi) - 1], \tag{1.1}$$

мы обычно ставим целью вычислить суперпропагатор

$$S(x) = [TL_{\mathrm{I}}(\phi_{\mathrm{in}}(x))\, L_{\mathrm{I}}(\phi_{\mathrm{in}}(0))]. \tag{1.2}$$

Формально

$$S(x) = \sum_{n=1}^{\infty} \frac{(\varkappa^2)^n}{n!} \langle T\phi_{\mathrm{in}}^n(x)\, \phi_{\mathrm{in}}^n(0)\rangle. \tag{1.3}$$

Мы ограничимся рассмотрением частиц нулевой массы, для которых

$$(T\phi_{\mathrm{in}}(x)\, \phi_{\mathrm{in}}(0)) = D(x) = -\frac{1}{x^2} \tag{1.4}$$

(фурье-образ пропорционален $1/p^2$). При вычислении члена вида $(T\phi^2(x)\phi^2(0))$ первая проблема заключается в том, какой смысл приписать оператору $\phi^2(x)$. В обычной процедуре используется равенство (1.4) для определения *нормального произведения*$:\phi^2$: из соотношения $\phi^2(x) = :\phi^2(x). + D(0)$. Здесь $D(0)$ — бесконечная перенормировочная константа $\lim\limits_{x\to 0} 1/x^2$. Теперь можно показать, что

$$(T:\phi^2(x)::\phi^2(0):) = \frac{2!}{(x^2)^2} \tag{1.4'}$$

с точностью до обобщенно-функциональной неоднозначности вида $b\delta^4(x)$ (b-неоднозначности). В этой простейшей ситуации уже содержатся три проблемы, составляющие сущность нашего обсуждения:

а) *Нормальное упорядочение*: содержится ли в $D(0)$ какая-либо физика, которую мы теряем при процедуре нормального упорядочения?

б) Величина $(1/x^2)^2$ есть произведение сингулярных распределений $(1/x_2) \otimes (1/x^2)$; существует ли какое-либо естественное определение фурье-образа этой величины?

в) Роль «константы неоднозначности» b.

Общепринятая теория перенормировок трактует проблемы (б) и (в) как части одной проблемы; обычно в фурье-пространстве записывают

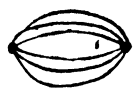
$$\int \frac{1}{x^2} \otimes \frac{1}{x^2} e^{ipx} d^4x = \frac{1}{(2\pi)^4} \int \frac{d^4k}{(p-k)^2 k^2}.$$

Интеграл в правой части (логарифмически) расходится. Для выделения из этого интеграла бесконечности разработана вычитательная процедура и постоянная b используется соответствующим образом для компенсации этой бесконечности.

Этот метод подавления бесконечности становится безнадежно сложным, когда мы рассматриваем объекты вида

$$\left(T : \phi^n_{\text{in}}(x) : : \phi^n_{\text{in}}(0) :\right),$$

которые представляют коконообразные графы с n линиями



содержащие $n-1$ расходящихся внутренних интегрирований в пространстве импульсов. В этом состоит одна из причин, по которой неперенормируемые теории с *полиномиальными* лагранжианами (например, вида $L_{\text{вз}} = g\phi^5$ или $g\phi^6$ и т. д.) были вскоре оставлены. Даже вычитательная процедура для подобных теорий трудно разрешима.

2. *Метод Гельфанда.* Теории, основанные на неполиномиальных лагранжианах, допускают еще один подход, основанный на процедуре Гельфанда — Шилова [1]. В этом подходе мы *разделяем проблемы (б) и (в)*. В основном это происходит потому, что суперпропагатор в подобных теориях представляет собой сумму ряда сингулярных функций:

$$S(x) = \sum \frac{1}{n!} \left(\frac{-\varkappa^2}{x^2} \right)^n.$$

Эта сумма гораздо менее сингулярна при $x \to 0$ по определенному направлению, чем каждый отдельный член ряда. (Грубо говоря, $\exp(-\varkappa^2/x^2) \to 0$, когда x^2 пространственноподобно и $\varkappa^2$ отрицательно. *Для остальных направлений* и остальных значений $\varkappa^2$ вопрос решается аналитическим продолжением.)

Для более точного рассмотрения вернемся к соотношению (1.4′). Мы хотим вычислить фурье-образ величины $(1/x^2)^2$, или в более общем случае величины $D^n(x) = (1/x^2)^n$. Гельфанд и Шилов заметили, что поскольку фурье-образ величины $(1/x^2)^z$

при $0 < \operatorname{Re} z < 2$ является хорошо определенным классическим математическим объектом и пропорционален величине

$$\frac{\Gamma(2-z)}{\Gamma(z)}\left(\frac{1}{p^2}\right)^{2-z},$$

фурье-образ величины $(1/x^2)^n$ при значении n, лежащем вне указанной области, может быть определен надлежащим *аналитическим продолжением этой функции по переменной* z. (Изящество этого определения контрастирует с неуклюжестью обычной вычислительной процедуры с ее многочисленными расходящимися внутренними интегрированиями. Термин «надлежащий» мы уточним немного ниже.)

Метод Гельфанда — Шилова рассматривался в физической литературе Гюттингером [2] в 1966 г., а в эквивалентной формулировке Густафсоном [3] даже еще раньше. Боллини и Джиамбьяджи [4] (см. также [5—7]), по-видимому, впервые использовали этот метод, чтобы переформулировать общепринятую теорию перенормировок. Однако сила и ценность этого метода становятся особенно очевидными, когда мы их используем для неполиномиальных лагранжианов, так как в этом случае несколько неясное понятие «надлежащее аналитическое продолжение» по переменной z становится согласованным с аналитическими свойствами суперпропагатора $S(x)$ по переменной $\varkappa^2 D(x)$.

Чтобы дать скелетную схему метода, рассмотрим суперпропагатор для лагранжиана

$$L_{\mathrm{I}} = (:g^2 e^{\varkappa\phi}: - 1).$$

Этот суперпропагатор представляет собой целую функцию от $\varkappa^2 D(x)$:

$$S(x) = g^2 \sum_{n=1}^{\infty} \frac{1}{n!}[\varkappa^2 D(x)]^n. \tag{1.5}$$

Во-первых, применим к суперпропагатору (1.5) преобразование Зоммерфельда — Ватсона:

$$S(x) = \frac{g^2}{2\pi i}\int \frac{dz}{\Gamma(z+1)}\frac{1}{\operatorname{tg}\pi z}[\varkappa^2 D(x)]^z. \tag{1.6}$$

Условия, при которых справедлив переход от (1.5) к (1.6), установлены в работах Волкова, Салама, Стрэтди, Лемана и Польмейера [8—13] [1]). При записи выражения (1.6) играет роль

[1]) О технике работы с неполиномиальными лагранжианами см. обзор [14].

b-неоднозначность, которую мы обсудим позднее. Контур интегрирования, как обычно, обходит положительную действительную ось от $\operatorname{Re} z < 1$ до бесконечности.

Во-вторых, повернем контур интегрирования так, чтобы он стал параллелен мнимой оси, — в данном конкретном случае вдоль $0 < \operatorname{Re} z < 1$. Условие Гельфанда — Шилова для «классического» фурье-преобразования удовлетворено, и мы можем написать

$$\tilde{S}(p) = \int\limits_{0 < \operatorname{Re} z < 1} (\varkappa^2)^z \left(\frac{1}{p^2}\right)^{2-z} \frac{\Gamma(2-z)\,dz}{\Gamma(z)\,\Gamma(z+1)\operatorname{tg}\pi z}. \qquad (1.7)$$

Подынтегральное выражение имеет полюс кратности один при $z = 1$ [соответствующий члену $-\varkappa^2/x^2$ в суперпропагаторе $S(x)$] и двойные полюса при $z = 2, 3, 4, \ldots$. Это порождает характеристический член, пропорциональный $(\varkappa^2 p^2)^r \log(\varkappa^2 p^2)$. *Эта логарифмическая зависимость функции Грина от малой константы связи является отличительным признаком теорий, основанных на неполиномиальных лагранжианах.*

3. *Подавление бесконечностей.* С целью убедиться в том, что указанная логарифмическая зависимость от малой константы связи означает подавление бесконечностей, рассмотрим смешанную теорию с лагранжианом

$$L_{\text{вз}} = g\chi^3 \exp(\varkappa\phi).$$

Суперпропагатор имеет вид

$$S(x) = g^2 \sum_{n=3} \frac{D^{n+3}(\varkappa^2)^n}{n!}.$$

Преобразование Зоммерфельда — Ватсона приводит к выражению

$$\frac{g^2}{2\pi i} \int\limits_{\operatorname{Re} z < 3} \frac{(\varkappa^2)^z D^{z+3}}{\operatorname{tg}\pi z\, \Gamma(z+1)}. \qquad (1.8)$$

Для удовлетворения условия Гельфанда — Шилова контур интегрирования должен быть сдвинут влево в область $\operatorname{Re} z + 3 < 2$. Это вполне возможно, поскольку подынтегральное выражение не сингулярно при $z = -1$. Фурье-образ $\tilde{S}(p)$ легко вычисляется и содержит члены, пропорциональные $g^2/\varkappa^2$, $g^2 \log(\varkappa^2 p^2)$, $g^2\varkappa^2 p^2 \log(p^2\varkappa^2)$ и т. д. Очевидно, член $g^2/\varkappa^2$ представляет собой проявление квадратичной бесконечности полиномиальной теории, задаваемой лагранжианом

$$g\chi^3 = \lim_{\varkappa \to 0} (g\chi^3)\, e^{\varkappa\phi}.$$

Подобным же образом член $\log(\varkappa^2 p^2)$ представляет собой проявление логарифмической бесконечности. Эти бесконечности обнаруживаются при $\varkappa \to 0$. *Другими словами, $\varkappa^{-1}$ является реалистическим регуляризующим обрезающим фактором в неполиномиальной теории, основанной на лагранжиане $L_{\text{вз}} = g\chi^3 \exp(\varkappa\phi)$.*

Б. Различие между конечными и перенормируемыми лагранжианами

Изложенным способом может быть регуляризован далеко не любой варинт появления бесконечности. Рассмотрим лагранжиан

$$L_{\text{II}} = e^{\varkappa\phi} - 1 - \varkappa\phi - \frac{\varkappa^2\phi^2}{2!}.$$

В данном случае суперпропагатор S_{II} имеет тот же вид, что и (1.6), но контур интегрирования лежит в области $2 < \operatorname{Re} z < 3$. Мы не можем поменять местами интегрирование по z и преобразование Фурье, поскольку условие Гельфанда — Шилова $\operatorname{Re} z < 2$ не удовлетворяется. Прежде чем это сделать, нужно записать

$$S_{\text{II}}(x) = S_{\text{I}}(x) - \frac{\varkappa^2 D^2}{2!}.$$

Суперпропагатор $S_{\text{I}}(x)$ можно подвергнуть преобразованию Фурье указанными выше методами, но член $(-\varkappa^2 D^2)/2!$ не поддается регуляризации этими методами (обозначим его НР). Он может быть регуляризован при помощи любого другого имеющегося в распоряжении метода; однако при этом нет основания для того, чтобы эффективное обрезание зависело от константы $\varkappa$. Лагранжиан L_{I} мы назовем *конечным*, а лагранжиан L_{II} — *перенормируемым*, поскольку последний нуждается в вычитательной константе обычного типа. В теории, основанной на лагранжиане L_{II}, по крайней мере одна физическая величина оказывается невычисляемой. В идеальной теории, разумеется, вообще не должно содержаться невычисляемых перенормировочных констант.

В. Неоднозначности

Обратимся теперь к проблеме неоднозначности [15]. Даже в *конечных* теориях (в смысле, указанном выше) возникает эта проблема. Источник этой проблемы лежит в *теории распределений*: распределение

$$(T : \phi^n(x) : : \phi^n(0) :)$$

неоднозначно с точностью до членов вида

$$\sum b_n (\partial^2)^{n-2} \delta(x).$$

С другой стороны, можно усмотреть эту b-неоднозначность и в формализме Зоммерфельда — Ватсона. При переходе от (1.5) к (1.6) и (1.7) экстраполированная функция была записана только на основании знания ее значения в целых точках. Для большей строгости нам следовало бы в выражение (1.6) вместо множителя $1/\operatorname{tg} \pi z$ включить более точный множитель

$$\frac{1}{\operatorname{tg} \pi z} + b(z).$$

Как я уже сказал, решающий вклад Лемана и Польмейера заключается в том, что они обнаружили существование простого физического критерия (применимого ко всем порядкам по главной константе связи g), используя который можно исключить из локализуемых теорий указанную обобщенно-функциональную неоднозначность. Основная идея заключается в том, что для *конечных* локализуемых теорий зависящий от b (b-неоднозначный) вклад в выражение (1.7) может быть отделен от вклада с $b = 0$ по характеру его поведения в p-пространстве при больших $|p^2|$. Зависящие от b члены не уменьшаются во всех направлениях в $|p^2|$-плоскости, в то время как не зависящие от b члены обладают этим свойством и, таким образом, определяют минимально сингулярный суперпропагатор $\tilde{S}(p)$. Тот же критерий был использован ранее Филипповым [16] для исключения неоднозначности в суперпропагаторе второго порядка, а также всеми исследователями, работающими в данной области. Леман и Польмейер дали общую формулировку проблемы, справедливую для суперпропагаторов всех высших порядков. Они показали далее, что это различие между b-зависимыми и b-независимыми членами не может быть проведено для полиномиальных теорий; *в этом отношении неполиномиальные теории намного совершеннее полиномиальных.*

Г. Нелокализуемые рациональные лагранжианы

Обратимся теперь к случаю *нормально упорядоченных* нелокализуемых теорий, например основанных на лагранжиане

$$L_{\mathrm{III}} = :\frac{g}{1 + \varkappa\phi}: .$$

В данном случае

$$S_{\mathrm{III}}(x) = g^2 \sum_{n=0}^{\infty} \Gamma(n+1) (\varkappa^2 D(x))^n. \tag{1.9}$$

Рассмотрим, чем отличается это выражение от суперпропагатора S_{I}. Рассматриваемый как ряд по величине $\varkappa^2 D(x)$, суперпропагатор (1.9) в противоположность (1.5) имеет *нулевой* радиус сходимости. Ефимов и Фрадкин [17—24] предложили *определять* суперпропагатор как борелеву сумму от (1.9). Формально можно записать

$$S_{\mathrm{III}}(x) = \sum_{n=0}^{\infty} \int_0^{\infty} e^{-\zeta} (\varkappa^2 D(x) \zeta)^n \, d\zeta, \tag{1.10}$$

$$S_{\mathrm{III}}(x) = \int_0^{\infty} \frac{1}{1 - \varkappa^2 D(x) \zeta} e^{-\zeta} \, d\zeta. \tag{1.11}$$

В качестве определения физического суперпропагатора Ефимов и Фрадкин приняли главное значение интеграла (1.11) в ζ-плоскости. Отправляясь от этого определения [или от эквивалентного выражения (1.10)], можно затем использовать методы Гельфанда — Шилова для фурье-преобразования в точности так же, как в случае (1.7).

Совершенно очевидно, что использование борелевой суммы вводит здесь новый дополнительный источник неоднозначности, *борелеву неоднозначность, не имеющую места в локализуемых теориях.* Одно из нежелательных следствий состоит в том, что предложенный Леманом и Польмейером прием обращения с b-неоднозначностями в данном случае бесполезен; как b-неоднозначные члены, так и члены с главным значением интеграла в фурье-образе выражения (1.11) спадают одинаково быстро в $|p^2|$-плоскости при больших $|p^2|$. Еще одну трудность отметил Фелс [27], показавший, что в этом случае не может быть гарантирована необходимая теоретико-полевая положительная определенность главного значения суперпропагатора (1.11).

Стрэтди, Дельбурго и автор доказали теорему, которая, возможно, исправит положение. Вызванные некоторыми замечаниями Фелса, наши соображения заключаются в следующем. *Трудности нелокализуемых теорий лежат в проблеме нормального упорядочения.* Если лагранжиан L_{III} не упорядочен нормально, а именно если мы формально оставляем в теории величину $D(0)$ как неопределенный параметр, то поведение суперпропагатора в *импульсном* пространстве в области высоких энергий коренным образом изменяется. Точнее, мы показали, что для рациональных нелокализуемых теорий [с лагранжианом вида $\Sigma a_i/(1+\varkappa_i\phi)$] не приведенные к нормальному виду ряды разложений для суперпропагаторов сходятся, если их рассматривать как ряды по $D(x)$. Радиус сходимости задается величиной $D(0)$.

Если $D(0) = \infty$, то радиус сходимости бесконечно велик, для определения (1.11) не требуется применения приема, связанного со взятием главного значения интеграла, и суперпропагаторы явным образом удовлетворяют критерию положительной определенности. Это очень приятно, но, к сожалению, как заметил Фрид [28], если положить

$$D(0) = \lim_{x^2 \to \infty} \left(-\frac{1}{x^2}\right) = \infty,$$

то почти все матричные элементы обращаются в нуль. Ясно, что величина $D(0)$ должна быть перенормирована. Если имеются физические основания считать, что величина $D(0)$ должна быть *конечной* [например, если в теории скалярное поле [1]) имеет ненулевое ожидаемое значение $\langle\phi\rangle \neq 0$, которое может быть выражено как функция от $D(0)$], то суперпропагаторы должны обладать *конечными* радиусами сходимости. В p-пространстве *не упорядоченных нормально рациональных теорий* двухточечная спектральная функция $\rho(p^2)$ должна вести себя как $\exp\sqrt{|p^2|}$, т. е. мы имеем *вполне локализуемую* ситуацию. Доказательство наших результатов просто и может быть легко воспроизведено. Для построения доказательства полезно отметить следующее. Если задан лагранжиан $L(\phi)$, то соответствующий нормально упорядоченный лагранжиан получается из него при помощи простого операторного тождества

$$:L(\phi): = \exp\left(\frac{D(0)}{2}\frac{\partial^2}{\partial\phi^2}\right) L(\phi).$$

Наш вывод состоит в том, что нормальное упорядочение неприменимо для нелокализуемых теорий [2]). *Рациональные* лагранжианы *вполне локализуемы* при условии, что $D(0)$ перенормируется до конечного ненулевого значения; они становятся *нелокализуемыми* только в том случае, если они нормально упорядочены [29, 30] (или, что эквивалентно, если $D(0)$ перенормируется до нулевого значения).

Мы еще не завершили исследования того, применим ли ко *вполне локализуемым* теориям прием Лемана и Польмейера изолирования b-неоднозначности, а также мы не можем сказать,

---

[1]) Например, для случая гравитации выражение $\langle g_{\mu\nu}\rangle$, связанное с космологической константой, можно выразить как функцию от $D(0)$. Я признателен проф. Гелл-Манну и проф. Зекслю за это замечание.

[2]) Уже давно было известно, что нормальное упорядочение неприменимо для калибровочных теорий. Это обусловлено тем, что при обыкновенном упорядочении трудно определить киральные и калибровочные преобразования и калибровочные свойства теории нарушаются.

обладают ли эти теории обрезающим фактором $\varkappa^{-1}$ или он видоизменяется вследствие появления новой константы $D(0)$. *Очевидно, что при применении нелокализуемых теорий величина $D(0)$ содержит некоторую физическую информацию.*

Д. Конечные нелокализуемые теории

Перечислим теперь те случаи, когда неполиномиальные теории *конечны*, а не только *перенормируемы* (т. е. нуждаются в конечном числе констант, не определенных и не вычисляемых в рамках данной теории). К сожалению, все выполненные до сих пор по этой проблеме работы (статьи Ефимова и Фрадкина, Дельбурго, Коллера и Салама [31], Тейлора и Кека [32]) концентрируют внимание на нормально упорядоченных рациональных лагранжианах вида

$$L = : g \frac{\phi^n}{(1 + \varkappa\phi)^w} : .$$

Приведем выводы, полученные в этих работах.

а) Лагранжиан L *конечен* при $n - w < 2$.

б) Лагранжиан L перенормируем при условии $2 \leqslant n - w < 4$, т. е. существует конечное число НР-членов, которые не могут быть вычислены через константы g и $\varkappa$, но для которых соответствующие бесконечности могут быть исключены перенормировкой физических констант. Отметим неожиданное неравенство $n - w < 4$. Можно было бы наивно ожидать, что теории будут перенормируемы при $n - w \leqslant 4$.

в) Теория *перенормируема* при всех $w > 0$ для смешанных теорий с лагранжианом вида

$$L = : g \frac{\bar{\psi}\psi\chi}{(1 + \varkappa\phi)^w} : , \tag{1.12}$$

в котором вес Дайсона полиномиальной части (т. е. $\bar{\psi}\psi\chi$) $\leqslant 4$ (вес Дайсона скалярного поля χ равен единице, вес Дайсона спинорного поля ψ равен $^3/_2$).

г) Мы полагаем, что если отказаться от нормального упорядочения (т. е. сохранять в теории величину $D(0)$ как неопределенную константу или выразить $D(0)$ через невычисляемую величину $\langle\phi\rangle$), то лагранжиан (1.12) приведет к конечной теории при условии $w > 2$. Для пограничного случая $w = 2$ единственные НР-члены — это те, которые соответствуют вакуум-вакуумным диаграммам.

д) Для экспоненциальных (локализуемых) лагранжианов ситуация более благоприятна в отношении *конечности* [33—36].

Лагранжианы вида $:g\phi^n e^{\varkappa\phi}:$ *конечны*. Этот результат тривиален для суперпропагаторов во втором порядке. Для суперпропагаторов более высокого порядка доказательство этого утверждения не встречает трудностей при $n \leqslant 2$. При $n > 2$ результат выводится дифференцированием $\exp \varkappa\phi$ по константе $\varkappa$.

е) Не столь легко для суперпропагаторов высшего порядка доказать конечность смешанного лагранжиана $:g\chi^n e^{\varkappa\phi}:$. При $n \leqslant 2$ доказательство достаточно просто. При $2 < n \leqslant 4$ мы полагаем, что рассматриваемые теории конечны, но полное детальное доказательство до сих пор еще не построено. Если рассматривать эффект сильного сглаживания бесконечностей, то, по-видимому, $e^{\varkappa\phi}$ ведет себя так же, как рациональный лагранжиан с произвольно большим весом w:

$$\exp \varkappa\phi = \lim_{w\to\infty} \frac{1}{(1-\varkappa\phi/w)^w}.$$

Е. Исчисление производных и обобщенное нормальное упорядочение

Обратимся, наконец, к лагранжианам, содержащим производные; они соответствуют большинству физически интересных случаев. В моем докладе на конференции в Майами в 1970 г. содержалось замечание, что, хотя на первый взгляд дифференцирование величины $1/x^2$ должно увеличивать ее сингулярность, на практике мы убеждаемся, что это увеличение не столь фатально, как при умножении на ту же величину, возведенную в соответствующую степень. В иной формулировке множитель вида $\partial_\mu\partial_\nu\phi$ в лагранжиане проявляется в отношении индуцирования бесконечностей гораздо слабее, чем множитель вида $\delta_{\mu\nu}\phi^3$, хотя на первый взгляд оба они обладают одним и тем же весом Дайсона ($w = 3$). Это наблюдение, выведенное эмпирически из вычислительной практики, имеет важные следствия в вопросе о различии между *конечностью* и *перенормируемостью* в лагранжианах кирального или гравитационного типа, содержащих производные. В этом случае свободный лагранжиан и лагранжиан взаимодействия существенно смешаны. В качестве простого примера рассмотрим лагранжиан

$$L_{\text{полн}} = \frac{(\partial\varphi)^2}{(1+\varkappa\varphi)^2} + m^2\varphi^2.$$

Приписывая каждой производной вес 1, получаем, что вес Дайсона всего лагранжиана $L_{\text{полн}}$ равен 2. Однако, когда свободный лагранжиан выделен из $L_{\text{полн}}$, а мы вынуждены сделать это перед тем, как начать вычисления по теории возмущений, вес

Дайсона остающегося лагранжиана взаимодействия $L_{\text{вз}}$ оказывается равным 4. Каков же истинный вес? *Конечна* или *перенормируема* теория? Другими словами, если мы запишем

$$L_{\text{вз}} = L_{\text{полн}} - (\partial\varphi)^2,$$

то будут ли две производные во втором члене действительно порождать новые НР-члены?

Чтобы ответить на эти вопросы, представляется существенным расширить исчисление обобщенных функций Гельфанда — Шилова с целью включения производных. Это может быть выполнено следующим образом. Равенство

$$\partial_\mu \left(\frac{1}{x^2}\right)^{z_1} \otimes \left(\frac{1}{x^2}\right)^{z_2} = \frac{z_1}{z_1 + z_2}\, \partial_\mu \left(\frac{1}{x^2}\right)^{z_1+z_2} \tag{1.13}$$

справедливо для всех x при $0 < \operatorname{Re} z_1, z_2, z_1 + z_2 < 2$. Если z_1 и z_2 лежат вне этой области, то две части равенства (1.13) могут различаться в точке $x = 0$ на члены, содержащие дельта-функцию $\delta(x)$, ее производные, а также множители, содержащие $[D(0)]^r$. Мы полагаем, что это равенство и его обобщения дают правильное расширение исчисления Гельфанда — Шилова на случай наличия производных. Для вычисления левой части (1.13) перепишем ее в виде, аналогичном правой части, и совершим надлежащее аналитическое продолжение из области $0 < \operatorname{Re} z_1, z_2, z_1 + z_2 < 2$ [1]). Важное свойство равенства (1.13) состоит в том, что число производных в обеих его частях сохраняется. В результате полная сингулярность от линий $(1/x^2)^{z_1+z_2}$ при дифференцировании не усиливается в одной части равенства (1.13) больше, чем в другой части. Другими словами, производные и сингулярные функции представляют собой несколько различные объекты с обобщенно-функциональной точки зрения (я недостаточный специалист по теории обобщенных функций, чтобы формализовать это утверждение), и это различие должно быть принято во внимание. Чтобы увидеть силу и смысл этого

[1]) Формулы (1.13) и (1.14) мы будем называть формулами «обобщенного нормального упорядочения» для ситуаций, содержащих производные. Мы полагаем, что использование этих формул исключает те величины $D(0)$, которые возникают при наложении производных [т. е. от членов вида $D(x)\partial^2 D(x)$]. Те величины $D(0)$, которые возникают от частей лагранжиана, не содержащих производных, могут быть, по-видимому, исключены при помощи стандартного нормального упорядочения, так что применение двух процедур *совместно*, возможно, даст полную схему исключения всех величин $D(0)$ для локализуемых калибровочных теорий. Это предположение требует дальнейшего исследования.

правила, рассмотрим простейшее обобщение равенства (1.13), задаваемое соотношением

$$\partial_\mu\left(\frac{1}{x^2}\right)^{z_1}\otimes\partial_\nu\left(\frac{1}{x^2}\right)^{z_2}=$$
$$=\frac{z_1z_2}{(z_1+z_2)(z_1+z_2+1)}\left(\partial_\mu\partial_\nu+\frac{1}{2(z_1+z_2-1)}\partial^2\delta_{\mu\nu}\right)\left(\frac{1}{x^2}\right)^{z_1+z_2}, \quad (1.14)$$
$$0<Re\, z_1z_2,\quad z_1+z_2<1.$$

Члены правой части этого равенства содержат две производные — ровно столько же, сколько члены левой части. Рассмотрим, как соотношение (1.14) используется при обычном вычислении собственной энергии фотона [будем рассматривать электроны нулевой массы с пропагаторами $S(x)=\eth(1/x^2)$]:

$$\Pi_{\mu\nu}(x)=\mathrm{Sp}\,\gamma_\mu\eth\left(\frac{1}{x^2}\right)\gamma_\nu\eth\left(\frac{1}{x^2}\right)=2(\partial_\mu\partial_\nu-\partial_{\mu\nu}\partial^2)\left(\frac{1}{x^2}\right)^2. \quad (1.15)$$

[Мы продолжили соотношение (1.14) за область его определения; эта операция, разумеется, всегда требует осторожности.] Выражение (1.15) правильно отражает поперечный характер собственной энергии фотона. Используя даже наиболее устарелый и несовершенный из способов вычисления, получаем, что фурье-образ величины $\Pi_{\mu\nu}(x)$ равен

$$2(p_\mu p_\nu-\delta_{\mu\nu}p^2)\int\frac{d^4k}{(p-k)^2k^2}.$$

В этом выражении не содержится никакого намека на порождение какой-либо собственной массы фотона. В обычном способе вычисления появление собственной массы у фотона является результатом беззастенчивой манипуляции, связанной с тем, что при этом забывается правило сохранения числа производных и $\Pi_{\mu\nu}(x)$ записывается в виде

$$\Pi_{\mu\nu}(x)=2\partial_\mu\partial_\nu\left(\frac{1}{x^2}\right)^2-16\delta_{\mu\nu}\left(\frac{1}{x^2}\right)^3. \quad (1.16)$$

Таким образом, в обычном подходе второй член в правой части выражения (1.15) $\partial^2(1/x^2)^2$ заменен на член $8/(x^2)^3$. Поскольку выражение $(1/x^2)^3$ более сингулярно, чем $(1/x^2)^2$, то ошибочно утверждается, что теория приводит к (квадратично бесконечной) собственной массе фотона.

Конечно, такое осторожное обращение с производными наиболее важно в тех случаях, когда они встречаются в вершинах, в которых имеются внешние линии. Но я обращаю внимание на этот вопрос потому, что считаю, что сила аналитических методов Гельфанда и их обобщения на связь через производные еще не оценена в должной мере и что половина наших трудностей в теории перенормировки обусловлена плохой математикой. Если принять, что при вычислении фурье-образов произведений распределений вида $(1/x^2)^{z_1} \otimes (1/x^2)^{z_2}$ не используются теоремы о свертке, то их не следует использовать [1] и для $\partial_\mu (1/x^2)^{z_1} \otimes \otimes \partial_\nu (1/x^2)^{z_2}$.

Ж. Поведение локализуемых теорий в области больших энергий на массовой оболочке

Еще одна проблема для неполиномиальных лагранжианов чисто теоретико-полевого аспекта заключается в следующем. Глазер, Мартин и Эпштейн [37] в своей фундаментальной работе показали, что во всех *локализуемых* теориях должна иметь место ограниченная в смысле Фруассара зависимость элементов S-матрицы в области высоких энергий. Этого результата нельзя достигнуть разложением Волкова — Лемана [8—12] по главной константе связи. В этом разложении любой заданный порядок ведет себя так же, как $\exp(|p^2|^\alpha)$, $\alpha < 1/2$. Я полагаю, что эта трудность будет преодолена, когда будет выполнено суммирование по главной константе связи. Мы уже знаем, что для обычных полиномиальных теорий этот тип суммирования — выполняемый либо путем прямого суммирования диаграмм, либо с использованием аппроксимаций Паде, либо при помощи неявного использования интегрального уравнения Бете — Солпитера — приводит к изменению поведения каждой отдельной диаграммы в области высоких энергий. Более конкретно мы знаем, что если даже отдельные диаграммы ведут себя как S^α (α — константа), то сумма лестничных диаграмм обнаруживает резкое изменение поведения и демонстрирует реджевское поведение $S^{\alpha(t)}$. Подобным же образом я полагаю, что сумма цепочки суперпропагаторов, вероятно, удовлетворяет результатам Глазера — Мартина — Эпштейна и обладает ограниченной по Фруассару зависимостью в области высоких энергий, хотя ни одного доказательства этого утверждения не построено. Указания,

---

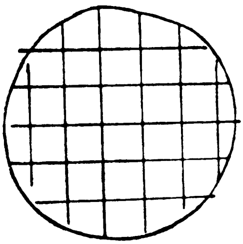
[1]) Следует подчеркнуть, что до сих пор не исследована ситуация вида

$$\frac{\partial}{\partial x_\mu}\left[\frac{1}{(x-y)^2}\right]^{z_1} \otimes \frac{\partial}{\partial x_\nu}\left[\frac{1}{(x-u)^2}\right]^{z_2}, \quad y \neq u.$$

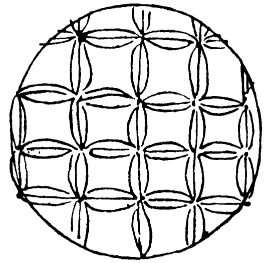
Это аналогично случаю Гельфанда — Шилова, где обычные методы свертки должны быть использованы для вычисления фурье-образа выражения

$$\left[\frac{1}{(x-y)^2}\right]^{z_1} \otimes \left[\frac{1}{(x-u)^2}\right]^{z_2}, \quad y \neq u.$$

подтверждающие изложенную точку зрения, возможно, содержатся в работах Олесона, Нильссона, Сускинда, Сакиты, Вирасоры и др. [38], в которых рассматриваются сеточные диаграммы типа фиг. 1. В этих работах показано, что если подобные диаграммы связаны с пропагатором вида $\exp(\varkappa^2 p^2)$, а не с пропагатором фейнмановского типа $1/p^2$, то сумма этих диаграмм демонстрирует венециано-реджевское (ограниченное по Фруассару) поведение в области высоких энергий. Но $\exp(\varkappa^2 p^2)$ представляет



Фиг. 1.



Фиг. 2.

собой как раз тот тип поведения в области высоких энергий, который мы получаем из суперпропагатора (в нелокализуемых теориях). Это наводит на мысль, что метод Венециано и метод неполиномиальных лагранжианов могли бы сблизиться благодаря диаграммам типа, показанного на фиг. 2, где каждый элемент сетки представляет собой суперпропагатор. Названные выше авторы показали, что член $(\varkappa^2)^{-1}$ в их формуле дает наклон траектории Редже. В этом заключается другая возможная физическая роль малой константы связи $\varkappa$. (Стоит подумать о том, чтó это означает для гравитации.)

ЧАСТЬ II

В этой части наших докладов д-р Стрэтди изложит опубликованную в Триесте работу, посвященную эйнштейновской теории гравитации и ее роли в подавлении бесконечностей лептонной электродинамики [39—41]. В своих замечаниях я особенно подчеркну три аспекта исследования, которое мы выполнили после опубликованной работы.

1. Мы теперь можем провести вычисления таким образом, чтобы сохранялась *калибровочная инвариантность*.

2. Теперь мы яснее представляем себе, почему именно *тензорная* (а не *скалярная*) гравитация ответственна за подавление бесконечностей. Это согласуется с физически интуитивным

убеждением (высказанным мне, в частности, проф. Вайскопфом), что подавление бесконечностей должно явиться следствием флуктуаций на световом конусе, специфичных для тензорной гравитации в ее (шварцшильдоподобном) метрическом аспекте.

3. Для демонстрации этого различия между скалярной и тензорной теориями гравитации мы нуждаемся в лучшем понимании роли теорем эквивалентности для преобразования полевых переменных. Поскольку, как мы видели в части I, локализуемые лагранжианы представляют собой вполне обычные и приемлемые типы лагранжианов и так как для операторов в локализуемых теориях выполняется микропричинность, мы можем воспользоваться результатами Борхерса и принять, что теоремы эквивалентности для S-матриц справедливы для локализуемых теорий. *Это приведет к необходимости использовать для гравитации не обычную рациональную, а экспоненциальную параметризацию.*

Как доказано в части I, п. Д, экспоненциальная неполиномиальность обладает сильно выраженной способностью сглаживать бесконечности. Это означает, что в теориях, учитывающих наличие гравитации, главная трудность состоит не в том, чтобы убедиться, могут ли быть регуляризованы бесконечности или нет, а в том, чтобы убедиться, что может быть сохранена калибровочная инвариантность.

А. Калибровочно-инвариантные вычисления в тензорной гравитации

Как покажет д-р Стрэтди, лагранжиан квантовой электродинамики, учитывающий наличие гравитации, может быть представлен в виде

$$L_{\text{полн}} = L_{\text{грав}} + \frac{L^{\mu a}\bar{\psi}\gamma_a(\nabla_\mu - ieA_\mu)\psi + m\bar{\psi}\psi}{\det L} + \frac{g^{\mu\mu'}g^{\nu\nu'}F_{\mu'\nu'}F_{\mu\nu}}{\det L}. \quad (2.1)$$

Здесь $L^{\mu a}$ — тетрадное гравитационное поле, которое в экспоненциальной параметризации записывается как

$$L^{\mu a} = \exp(\varkappa\beta_{\lambda c}h^{\lambda c})$$

[$\beta_{\lambda c}$ — антисимметричные матрицы 4×4, а $h^{\lambda c}$ — физическое (квантованное) поле, описывающее тензорные кванты]. Оператор ∇_μ в (2.1) обозначает ковариантное дифференцирование; $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, в то время как эйнштейновское поле $g^{\mu\nu}$ равно билинейному произведению тетрадных полей $L^{\mu a}L^{\nu}_{a}$. Отметим, что

$$\det L = \exp[\operatorname{Sp}(\varkappa h)].$$

Простота этого выражения делает работу с экспоненциальной параметризацией (поскольку мы имеем дело с $\det L$) даже более простой, чем с рациональной параметризацией, использованной нами в предыдущих работах. Соответствующий лагранжиан для случая скалярной гравитации может быть получен из (2.1) подстановкой

$$\begin{aligned} &L^{\mu a} = \exp(\varkappa\phi)\,\eta^{\mu a}, \qquad \eta^{\mu a} = \mathrm{diag}\,(1, -1, -1, -1), \\ &g^{\mu\nu} = \exp(\varkappa\phi)\,\eta^{\mu\nu}, \quad \det L = \exp(4\varkappa\phi); \end{aligned} \qquad (2.2)$$

при этом полный лагранжиан $L_{\text{полн}}$ принимает вид

$$\begin{aligned} L_{\text{полн}} = L_{\text{грав}} + [\bar{\psi}\gamma_\mu(\partial_\mu - ie_0A_\mu)\,\psi]\exp(-3\varkappa\phi) + \\ + m_0\bar{\psi}\psi\exp(-4\varkappa\phi) + F_{\mu\nu}F_{\mu\nu}. \end{aligned} \qquad (2.3а)$$

Заметим, что, как и можно было ожидать из физических соображений, *фотонное поле не взаимодействует со скалярной гравитацией.*

Совершим теперь преобразование поля

$$\psi' = \exp\left(-\frac{3}{2}\varkappa\phi\right)\psi. \qquad (2.3б)$$

При этом взаимодействие с полем $\varkappa\phi$ полностью исчезнет из всех членов лагранжиана, за исключением члена с массой электрона. В пределе $m_0 = 0$ скалярная гравитация и (безмассовые) электроны также не взаимодействуют.

Но если бы δm и δe были строго определенными физическими величинами на массовой оболочке, то можно было бы без колебаний сказать, что, подобно всем элементам S-матрицы на массовой оболочке, величины δm и δe, вычисляемые из лагранжиана (2.3а), не могут содержать константы $\varkappa$ в пределе $m_0 = 0$. *Поэтому скалярная гравитация не может играть регуляризующей роли* для безмассовых электронов [1]). Для электро-

[1]) Могут быть две причины подвергать сомнению это утверждение. *Первая*: может ли изменить ситуацию включение массового члена? *Вторая*: поскольку δm и δe в точной теории выражены как интегралы от спектральных функций, применимы ли в действительности теоремы эквивалентности? Мое личное мнение таково, что они применимы. Мы знаем, например, что как δm, так и δe обладают вместе со строгими значениями величин на массовой оболочке свойством калибровочной инвариантности (в отличие от z_2). Однако единственный способ проверить это заключается в проведении вычислений. К сожалению, для прямой проверки теорем эквивалентности требуется суммирование бесконечных цепочек диаграмм. [Вспомните усилия, которые потребовались для доказательства теорем эквивалентности с использованием разложений теории возмущений для лагранжианов Дайсона — Нельсона $\bar{\psi}\gamma_\mu\gamma_5\partial_\mu\phi\psi$ и $\bar{\psi}'\exp(i\gamma_5\phi)\,\psi'$.] Проблема еще более сложна из-за необходимости сохранять калибровочную инвариантность на каждой стадии вычислений и успешно справиться с проблемой нормального упорядочения (ср. примечание на стр. 195).

нов, обладающих массой, ситуация, возможно, лучше. Однако до сих пор мы не могли показать, что массовый член в (2.3а) достаточен для регуляризации всех бесконечностей.

К счастью, истинная гравитация носит тензорный характер, и хотя преобразование

$$\psi' = (\det L)^{-1/2}\,\psi \tag{2.4}$$

устраняет[1]) множитель $(\det L)^{-1}$ из члена кинетической энергии и члена электрического тока электронного лагранжиана, все же взаимодействие с гравитацией имеет место как для электрона, так и (что более важно) для фотона. Для симметрии совершим также преобразование

$$A'_\mu = (\det L)^{-1/2}\,A_\mu. \tag{2.5}$$

Возникающий при этом лагранжиан имеет вид

$$L_{\text{полн}} = L_{\text{грав}} + L^{\mu a}\bar{\psi}'\gamma_a\left[\nabla_\mu - ieA'_\mu\,(\det L)^{1/2}\right]\psi' + \\ + m_0\bar{\psi}'\psi' + g^{\mu\mu'}g^{\nu\nu'}F'_{\mu\nu}F'_{\mu'\nu'}. \tag{2.6}$$

Здесь

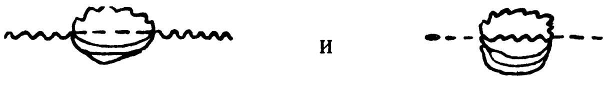
$$F'_{\mu\nu} = \left(\partial_\mu A'_\nu - \partial_\nu A'_\mu + A'_\nu\partial_\mu \log\sqrt{\det L} - A'_\mu\partial_\nu \log\sqrt{\det L}\right). \tag{2.7}$$

Преобразования (2.4) и (2.5) назовем *стандартными* преобразованиями, а лагранжиан (2.6) — стандартной формой лагранжиана. Преимущество стандартной формы заключается в том, что содержащиеся в (2.6) члены свободной кинетической энергии не содержат в качестве множителей $\det L$ в некоторой степени.

В наших ранее опубликованных работах мы эффективно рассмотрели те диаграммы для собственной энергии электрона и фотона, которые задаются итерацией второго порядка от лагранжиана

$$L_{\text{I}} = -ieL^{\mu a}\bar{\psi}'\gamma_\mu\psi' A'_\mu(\det L)^{1/2}, \quad \det L = \exp(\operatorname{Sp}\varkappa h). \tag{2.8}$$

Эти диаграммы для собственных энергий электрона и фотона имеют вид соответственно



Здесь волнистая линия соответствует электрону, штриховая — фотону и сплошная — гравитону. Ясно, что для сохранения

[1]) Отметим, что преобразование (2.4) не совпадает с преобразованием (2.3б), где мы положили $\psi' = (\det L)^{-3/8}\psi$.

калибровочной инвариантности существенно включение дальнейших диаграмм, возникающих из фотон-гравитонных членов:

$$L_{II} = (g^{\mu\mu'}g^{\nu\nu'} - \eta^{\mu\mu'}\eta^{\nu\nu'})\, F'_{\mu\nu}F'_{\mu'\nu'}. \tag{2.9}$$

В опубликованных работах мы не вычислили эти диаграммы (вследствие встреченных трудностей), и результат явно не был калибровочно-инвариантен.

Теперь оказывается возможным незначительным изменением наших прежних методов сохранить калибровочную инвариантность и в то же время гарантировать подавление бесконечностей [с точно такими же, как мы проверили, результатами, какие были получены ранее, если учитываются члены, содержащие ($\alpha \log \varkappa^2 m^2$), в вычисленных выражениях для δm и δe].

Произведем преобразование электронного поля (2.4), а фотонное поле оставим без изменения. Тогда получим

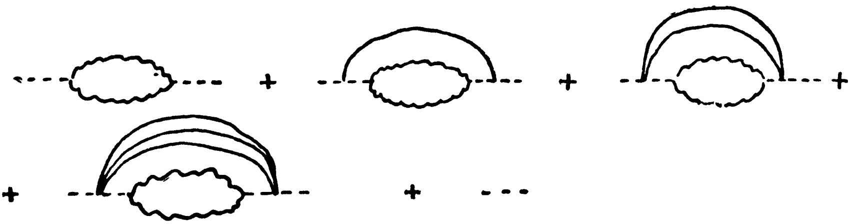
$$L_{\text{полн}} = L_{\text{грав}} + L^{\mu a}\bar{\psi}'\gamma_a (\nabla_\mu - ieA_\mu)\,\psi' + \\ + m_0\bar{\psi}'\psi' + g^{\mu\mu'}g^{\nu\nu'}F_{\mu\nu}F_{\mu'\nu'}\,(\det L)^{-1}. \tag{2.10}$$

Чтобы прояснить механизм калибровочно-инвариантного подавления бесконечностей, возьмем приближенное выражение для (2.10). (Ниже мы оценим степень этого приближения.) Отбросим зависящие от $\varkappa$ члены в выражениях для $L^{\mu a}$ и $g^{\mu\nu}$ и в выражении для ковариантной производной ∇_μ (т. е. заменим эти выражения на $\eta^{\mu a}$, $\eta^{\mu\nu}$ и ∂_μ соответственно). Компоненты гравитационного поля h^{11}, h^{22}, h^{33} и h^{44}, конечно, остаются в выражении $\det L = \exp(\varkappa h^{aa})$. В результате мы получим приближенное выражение для лагранжиана (2.10), имеющее вид

$$L_{\text{полн}} = L_{\text{грав}} + \bar{\psi}'\gamma_\mu (\partial_\mu - ieA_\mu)\,\psi' + \\ + F_{\mu\nu}F_{\mu\nu}\exp(\varkappa h^{aa}) + m_0\bar{\psi}'\psi'. \tag{2.11}$$

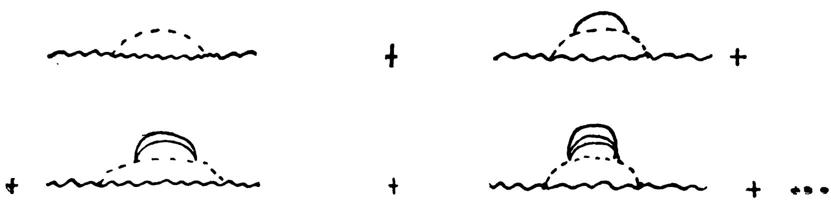
В теории, основанной на лагранжиане (2.11), по существу имеются четыре поля (h^{aa}), не взаимодействующие с электроном, но взаимодействующие калибровочно-инвариантным образом с *фотоном*. Далее, ток электронного заряда $\bar{\psi}'\gamma_\mu\psi'$ сохраняется, когда осуществляется взаимодействие как операторов ψ', так и гейзенберговских операторов. Это решающим образом отличается от случая лагранжиана (2.6). Именно это обстоятельство приводит к сохранению калибровочной инвариантности на всех этапах и в то же время гарантирует подавление бесконечностей, предоставляя требуемую *неполиномиальность* модифицированным фотонным членом в лагранжиане (2.11).

Диаграммы собственной энергии фотона, порождаемые лагранжианом (2.11) в первом порядке по e^2 и во всех порядках по $\varkappa$, показаны на фиг. 3, а диаграммы собственной энергии электрона — на фиг. 4. (Существуют также не относящиеся к делу модифицированные диаграммы с внешними фотонными линиями, которые мы не изобразили на фиг. 3.)



Фиг. 3. Диаграммы собственной энергии фотона.

В отдельной работе будет показано, что метод суммирования в неполиномиальном случае прекрасно работает для изображенных на фиг. 3 и 4 цепочек диаграмм. Решающий шаг



Фиг. 4. Диаграммы собственной энергии электрона.

здесь состоит в том, чтобы учесть то обстоятельство, что диаграмма для НР-члена (НР-диаграмма)

тождественно равна диаграмме

где крестики означают операцию включения члена $F_{\mu\nu}F_{\mu\nu}$. Таким образом, НР-диаграммы

 и 

в самом общем случае являются составными частями диаграмм, показанных на фиг. 3 и 4. Результаты *калибровочно-инвариантны*, в то время как численно $\delta m/m$, например, в точности равно (в согласии с тем, что утверждают теоремы эквивалентности) ранее полученному значению $(3/4\pi)\alpha \log \varkappa^2 m^2$ (с точностью до членов порядка α, $\alpha\varkappa^2$, ...). Можно показать, что эффект от включения членов, исключенных в процессе получения приближенного лагранжиана (2.11) (т. е. при замене $L^{\mu a}$ на $\eta^{\mu a}$ и т. д.), сведется к изменению членов только этих высших порядков α, $\alpha\varkappa^2$, ... и не окажет влияния на главные члены. Доказательство этого утверждения просто, но здесь мы его не приведем.

Б. Тензорная гравитация и искривленное пространство-время

Как я уже говорил, Вайскопф с его физической интуицией заметил ранее (в частном сообщении), что именно множитель $(\det L)^{-1}$, модифицирующий фотонный лагранжиан, главным образом ответствен за подавление бесконечностей, даже более, чем множитель $(\det L)^{-1}$, связанный с электронным лагранжианом [1]). Физически этого следует ожидать по той причине, что, как полагает Вайскопф, подавление бесконечностей в электродинамике связано с тем обстоятельством, что частота стоячих фотонных волн в искривленном пространстве-времени вокруг электрона имеет верхний предел $k_{\text{макс}} = 1/(\varkappa^2 m)$. Именно этот обрезающий фактор по частоте (виртуальных) фотонов, согласно мнению Вайскопфа, в действительности приводит к обрезанию в квантовой электродинамике.

В декабре прошлого года проф. Шифф формализовал для меня аргументы Вайскопфа. Поле Шварцшильда вокруг элек-

---

[1]) Говоря на языке математики, множитель $(\det L)^{-1}$ для электрона может быть исключен преобразованием (2.4), но это оказывается невозможным проделать для фотона. Сказанное представляет собой, конечно, сверхупрощение ситуации в случае тензорной гравитации, где неполиномиальность, связанная с величинами $L^{\mu a}$ (в явно экспоненциальной параметризации), способна сама по себе [даже без множителя $(\det L)$] подавлять бесконечности. Важно еще раз подчеркнуть, что независимо от того, какие цепи диаграмм суммируются, можно показать, что конечное регуляризованное значение для $\delta m/m$ и $\delta e/e$ не меняется с точностью до членов порядка $\alpha \log \varkappa^2 m^2$. Другими словами, неполиномиальности от различных членов не дают аддитивных эффектов при регуляризации бесконечностей в основных HP-диаграммах

 и 

трона задается метрикой

$$(ds)^2 = (dt)^2\left(1 - \frac{2mG}{r}\right) - \frac{(dr)^2}{1 - 2mG/r} - r^2\left[(d\theta)^2 + \sin^2\theta\,(d\phi)^2\right] \qquad (G = 8\pi\varkappa^2)$$

Если $\theta = \pi/2$ и $u = 1/r$, то траектория луча света дается уравнением

$$\frac{d^2u}{d\phi^2} + u = 3mGu^2.$$

При $u = \mathrm{const}$ и $u = u_0 = 1/3mG$ мы получаем круговую траекторию, которую луч проходит за время

$$t = t_0 = \frac{2\pi r_0}{\sqrt{1 - 2mG/r_0}} = 6\sqrt{3}\,\pi mG.$$

Таким образом, собственная угловая частота орбитального движения фотона по круговой орбите есть $1/(3mG)$, т. е. максимальная частота Вайскопфа равна $k_{\text{макс}} = 1/(3mG)$. Если это значение $k_{\text{макс}}$ подставить в старую формулу Вайскопфа

$$\frac{\delta m}{m} \approx \alpha \int\limits_0^{k_{\text{макс}}} \frac{d^3k}{|k^2|},$$

то мы получим наш результат для $\delta m/m$.

В. Связь между постоянной тонкой структуры и ньютоновской гравитационной постоянной

Сделаем одно заключительное замечание по поводу формулы

$$\frac{\delta m}{m} \approx \frac{3}{4\pi}\,\alpha \log \varkappa^2 m^2. \tag{2.12}$$

Этот член является первым в разложении, которое, согласно работе Вайскопфа 1939 г., должно иметь вид

$$\frac{\delta m}{m} \approx \sum a_n (\alpha \log \varkappa^2 m^2)^r, \quad r \leqslant n, \tag{2.13}$$

где включены высшие порядки по α[1]). Важный момент заключается в том, что эффективный параметр этого разложения

[1]) Первин [44] проверил, что в четвертом порядке по e регуляризация бесконечностей происходит ожидаемым образом, т. е. величина $(\alpha \log \infty)^2$ регуляризуется гравитацией до величины $(\alpha \log \varkappa^2 m^2)^2$.

есть $(\alpha \log \varkappa^2 m^2)$. Эта величина имеет удивительное значение, близкое к единице ($\alpha \log \varkappa^2 m^2 \approx 100/137$). В более ранних работах мы высказали предположение, что природа, вероятно, имела в виду, чтобы формулы (2.12) и (2.13) читались в обратном направлении, т. е. мы можем начать с допущения о том, что вся (или почти вся) собственная масса электрона имеет электромагнитную природу, так что $\delta m/m \approx 1$. На этом основании можно в обратном порядке определить величину $\alpha \log (\varkappa^2 m^2)$, что, возможно, поможет понять, почему это число эмпирически так близко к единице.

Проблема бесконечностей в электродинамике возникла вместе с классической электронной теорией Лоренца почти семьдесят лет назад. Уоллер в 1930 г. исследовал эти бесконечности, используя одночастичную теорию Дирака, и нашел, что они продолжают существовать даже после квантования. Современная формулировка проблемы датируется известной работой Вайскопфа 1934 г. Могут быть другие решения проблемы бесконечностей; например, было высказано предположение, что суммирование по главной константе связи α регуляризует бесконечности или что при введении нарушающей унитарность индефинитной метрики бесконечности исчезают. Возможно, это действительно так. Мы хотели бы лишь подчеркнуть, что в природе существует реальный мощный регуляризующий эффект тензорной гравитации, что этот эффект не мал ($\approx \log \varkappa^2 m^2 \approx \alpha^{-1}$) и регуляризация такова, что мы наглядно обнаруживаем все старые бесконечности, когда производим предельный переход при $\varkappa \to 0$. Первые опубликованные нами результаты вычислений были конечны, но калибровочно-неинвариантны. Мы работали с некоторой специальной *нелокализуемой* параметризацией гравитации (в которой тензор $g^{\mu\nu}$ рассматривается как основное поле, а тензор $g_{\mu\nu}$ — как неполиномиальная вспомогательная величина). Мы не понимали природы неоднозначностей — ни обобщенно-функциональной, ни борелевой. Теперь все это изменено. Используя приемы Лемана и его работу и применяя *локализуемую* (экспоненциальную) параметризацию гравитации, мы обращаемся с теорией поля, как с теорией, не представляющей ничего таинственного, — скорее ее можно уже назвать ортодоксальной. Возможность производить любые преобразования поля позволила нам продемонстрировать в явной форме калибровочную инвариантность взаимодействующей с гравитацией *конечной* электродинамики. Мы показали, что по крайней мере одно полное решение имеющей очень долгую историю проблемы бесконечностей теперь существует в рамках почти ортодоксальной теории и это решение может быть принято к сведению.

ЧАСТЬ III

А. Двутензорная теория гравитации

Поскольку $\varkappa_g^{-1} \approx 10^{18}$ БэВ, маловероятно, чтобы обрезание, возникающее из эйнштейновской гравитации, имело значение в физике сильных взаимодействий. Мы нуждаемся в универсальном неполиномиальном сильном взаимодействии (аналогичном гравитационному), которое может приводить к наличию обрезающего фактора порядка нуклонной массы. Эта часть доклада будет посвящена исследованию гипотезы о том, что эйнштейновская теория гравитации сама по себе может быть видоизменена таким образом, что лептоны будут взаимодействовать непосредственно с тензором $g^{\mu\nu}$, в то время как новый тензор $f^{\mu\nu}$ (представляющий массивный унитарный синглет 2^+) сильно взаимодействует с адронным тензором энергии с константой связи $\varkappa_f$. Для лагранжиана f-частиц мы принимаем эйнштейновскую форму. Постулируется ковариантный член, смешивающий тензорные поля f и g и гарантирующий, что одна комбинация $\bar{f} = f - g$ этих двух тензоров представляет массивные 2^+-частицы (с массой m_f), в то время как другая комбинация

$$\tilde{g} = \frac{\varkappa_g^2 f + \varkappa_f^2 g}{\varkappa_g^2 + \varkappa_f^2}$$

описывает безмассовые гравитоны.

Доклад д-ра Айшема будет посвящен введению в суть дела. Как я сказал, главное основание для постулирования этой особой теории описания явлений сильных взаимодействий заключается в получении из нее обрезания порядка $(\varkappa_f)^{-1} \approx m_f \approx$ ≈ 1 БэВ с целью регуляризовать бесконечности физики сильных взаимодействий. Однако поскольку F-доминантность тензора энергии сильных взаимодействий также может быть введена в теорию, представляется, разумеется, уместным изучить эту специальную гипотезу[1]. Реннер, Джонс и Раман сделают сообщения по этому вопросу. Как обычно в физике сильных взаимодействий, их заключение оказывается следующим:

[1]) Следует подчеркнуть, что не существует неотвратимой логической необходимости ни для интерполирования гейзенберговского поля F, чтобы получить связанную с ним физическую частицу, ни для утверждения о том, что полюс, соответствующий подобной частице, должен полностью доминировать в тензоре энергии-импульса, хотя и было бы лучше, если бы это выполнялось.

«В гипотезе нет ничего неправильного, положение вещей оптимистично, но не забывайте, что в физике сильных взаимодействий нет ничего окончательно ясного». Я хочу сделать только одно замечание в связи с сообщением д-ра Джонса. Выдвигая гипотезу о том, что рассматриваемые в двутензорной теории F-мезоны все еще не открыты и, возможно, лежат на траектории Померанчука (с массой около 1700—2000 МэВ, приписываемой в настоящее время наклону траектории Померанчука), мы попытались понять сохранение спиральности в s-канале (эмпирически связываемое с обменом померанчоном) как следствие того, что траектория Померанчука обладает связью через тензор энергии со своим 2^+-повторением. Нам удалось доказать существование тесной связи между этими двумя идеями; однако соответствующее выражение для адронного тензора энергии должно быть «минимальным». (Минимальный тензор энергии по существу является тензором энергии свободных частиц.)

Сказанное соответствует тому, что говорил Гелл-Манн в своем докладе на данной конференции. По-видимому, изучение энергетического поведения на близких расстояниях приводит к положению, утверждающему, что физика определяется полевыми операторами свободных частиц (в противоположность S-матричным элементам для свободных частиц), в смысле алгебры токов, реджевской связи или поведения на близких расстояниях. Это положение Гелл-Манна в применении к нашей ситуации приводит непосредственно к «минимальному» тензору энергии, как соответствующему тензору, возникающему в нашем лагранжиане и связанному с F-мезонами.

Б. Черные дыры в F-гравитационном поле

Хотя данная теория создавалась только для целей подавления бесконечностей, мне бы хотелось серьезно рассматривать F-гравитацию как модификацию обычной гравитации на близких расстояниях. Доклад проф. Дезера посвящен этому вопросу. Мы приветствуем заверения этого видного представителя небольшой избранной группы ученых, удачно сочетающих интерес к теории гравитации и к физике элементарных частиц, о том, что двутензорная гравитационная теория не содержит грубых ошибок (подобно нарушению принципа эквивалентности) в описании обычных гравитационных явлений. Однако эта теория обладает своими довольно странными особенностями.

1. Если мы рассматриваем две адронные частицы, то ясно (вследствие f — g-смешивания), что в линейном приближении

статический потенциал между ними будет задаваться выражением

$$V_{hh}(r) \sim -\frac{\varkappa_f^2}{\varkappa_f^2+\varkappa_g^2}\left[\varkappa_f^2\frac{e^{-m_f r}}{r}+\varkappa_g^2\frac{1}{r}\right].$$

Соответствующий потенциал между двумя лептонами имеет вид

$$V_{ll}(r) \sim -\frac{\varkappa_g^2}{\varkappa_f^2+\varkappa_g^2}\left[\varkappa_g^2\frac{e^{-m_f r}}{r}+\varkappa_f^2\frac{1}{r}\right].$$

Здесь, конечно, нет еще ничего удивительного. Удивительное появляется в адрон-лептонном гравитационном потенциале на расстояниях $\approx 10^{-13}$ см. Этот потенциал задается выражением

$$V_{lh} \sim -\frac{\varkappa_f^2\varkappa_g^2}{\varkappa_f^2+\varkappa_g^2}\left[-\frac{e^{-m_f r}}{r}+\frac{1}{r}\right].$$

Заметим, что сингулярность типа $1/r$ исчезла. Уровни лептонной и адронной материи при сближении на расстояние меньше 10^{-13} см не притягиваются (гравитационно) так интенсивно, как это предполагалось; чередующиеся уровни могли бы привести к частичной экранировке обычной гравитации.

2. Можно ожидать, что нестатический гравитационный потенциал между адронами V_{hh} будет содержать короткодействующую *отталкивательную* компоненту вследствие обменов f-гравитонами *спина 2* во втором и более высоких порядках. До сих пор мы не подтвердили вычислениями наличие этой *отталкивательной гравитации* и высоко оценили бы помощь в выяснении той ситуации, которая имеет место в высших порядках по F-мезон-нуклонной константе связи.

3. Во взаимодействии обычных высокочастотных гравитонов с большими концентрациями вещества можно ожидать обнаружения поверхностных эффектов, известных в аналогичной теории $\rho-\gamma$-смешивания адронной электродинамики. Высокочастотные g-гравитоны благодаря наличию члена со смешиванием могли бы превращаться в f-гравитоны, которые вследствие короткодействующего характера их силы поглощались бы главным образом на *поверхности* (большой) массы материи, а не проникали бы на внутренние уровни. Очень грубо говоря, для гравитационных эффектов можно ожидать зависимости от массы скорее типа $M^{2/3}$, чем M, где M — масса большого объекта. Это ослабление гравитации может оказаться важным с точки зрения наступления гравитационного коллапса. (Для аналогичного $\rho-\gamma$-смешивания оказывается, что фотоны с энергией

около 20 БэВ эмпирически обнаруживают зависимость типа $Z^{0,9}$ при взаимодействии с массивными ядрами заряда Z.)

4. До сих пор мы говорили об обычном коллапсе. Рассмотрим теперь F-гравитацию в ее «метрическом» аспекте. Для областей внутри адронной материи, где g-гравитация (в хорошем приближении), по-видимому, может не учитываться, мы имеем дело с пространством-временем «сильной» кривизны. Подойдем серьезно к этому аспекту F-гравитации. Обычная формула Шварцшильда для радиуса эйнштейновской гравитации нуждается, разумеется, в ревизии для случая f — g-смешивания, и я укажу ниже некоторые возможные модификации [1]. Однако, принимая формулы такими, какие они есть, мы можем оценить величину константы связи $\varkappa_f$ по аналогии с известной эддингтоновской оценкой величины константы связи $\varkappa_g$. Эддингтон принял, что вселенная представляет собой сферу Шварцшильда (внутри которой мы живем и вне которой ничего с нашей стороны не проникает). Радиус вселенной R_u, ее масса M_u и гравитационная константа $\varkappa_g$ $\left(G_g = 8\pi\varkappa_g^2\right)$ связаны соотношением

$$R_u \approx M_u G_g. \tag{3.1}$$

С той точностью, с которой известны величины R_u и M_u, константа G_g (и $\varkappa_g^2$), определяемая из этой формулы $\left(\varkappa_g^2 \approx 10^{-44} m_e^{-2}\right)$, оказывается грубо правильной [2].

---

[1]) Мое предположение состоит в том, что уравнение f — g-связи, будучи решенным в статическом случае, вместо характерного выражения для метрики Шварцшильда

$$g_{rr} = \frac{1}{1 - \dfrac{2MG_g}{r}}$$

даст что-то более похожее на

$$f_{rr} = \frac{1}{1 - 8\pi M\left(\dfrac{\varkappa_g^2}{r} + \dfrac{\varkappa_f^2 e^{-m_f r}}{r}\right)}$$

для источника в виде адронной частицы массы M, действующего на адронную пробную частицу. Это даст для обобщенного радиуса Шварцшильда для сингулярности F-поля формулу типа

$$R \approx \frac{1}{m_f} \log\left(\varkappa_f^2 M m_f\right).$$

Отметим слабую зависимость R от массы источника M. Даже при использовании этой формулы величина $\varkappa_f^2 m_f^2$ оказывается правильного порядка, причем типичные адроны рассматриваются как почти сколлапсированные объекты, для которых горизонты антисобытий достигают горизонтов событий.

[2]) Считаю приятным долгом выразить свою признательноость проф. Чандрасекару за беседы по этому вопросу.

Предположим теперь, что адроны представляют собой (почти) *сколлапсированные объекты в F-гравитационном поле.* Типичным адроном является сам F-мезон. Записывая $R \approx 1/m_f$ в (3.1), мы получаем соотношение вида

$$\frac{1}{m_f} \approx m_f \varkappa_f^2, \quad \text{т. е.} \quad \varkappa_f^2 \approx m_f^{-2},$$

которое опять дает грубо правильное значение константы сильной связи F-мезонов (см. примечание на стр. 210).

Я хочу сказать, что, если даже физика черных дыр не окажется пригодной для макровселенной, она будет иметь применения в микровселенной адронной материи. Некоторые из вас, быть может, знакомы с жаргоном физики черных дыр, например с понятиями «горизонт событий», «ловушечная поверхность», «горизонт антисобытий» (который в некоторых случаях может уничтожать горизонт событий и давать то, что я назвал почти сколлапсированной черной дырой). Одна из наиболее красивых формул, которые были недавно получены, утверждает, что условие для ловушечной поверхности (и в конечном счете для черной дыры), образованной объектом массы M с угловым моментом L и зарядом Q, дается неравенством

$$M^2\varkappa^2 > \frac{L^2}{M^2\varkappa^2} + Q^2.$$

Для электрона правая часть равна 10^{-44}, а левая часть — 10^{+44}. Поэтому нет опасности, что лептоны могут когда-нибудь сколлапсировать. Но рассмотрим теперь типичный адрон в *F-гравитационном* поле. В этом случае обе стороны неравенства порядка единицы. Заменим, как это всегда делают в физике элементарных частиц, заряд Q^2 на I-спин $I(I+1)$ или унитарно симметричные операторы Казимира, и вы увидите, как из F-гравитации может возникнуть известная нам массовая формула

$$M^4\varkappa^4 = j(j+1) + \varkappa^2 M^2 I(I+1).$$

В этой проблематике заключено невыявленное еще богатство и возможность создания нового языка для физики элементарных частиц.

ВЫВОДЫ

1. Лучший результат этой конференции заключается в нашей уверенности, что локализуемые неполиномиальные теории являются вполне приемлемыми теориями поля, в действительности более или менее ортодоксальными и даже скорее банальными с аксиоматической точки зрения. Но не только это: такие теории имеют то преимущество по сравнению с полиномиальными теориями, что неполиномиальные теории позволяют (при помощи

приема Лемана) производить вычисления однозначных и конечных значений перенормировочных констант.

2. Слишком долгое время специалисты по теории элементарных частиц игнорировали гравитацию. Нашим извинением может служить малость константы связи с гравитацией. Теперь мы поняли, что это пренебрежение гравитацией искажает пространство-время и является одним из источников возникновения обычных бесконечностей в квантовой электродинамике. Логарифмическая бесконечность, например, есть просто результат того, что мы полагаем $\varkappa_g = 0$ в выражении $\alpha \log \varkappa_g^2 m^2$. Поскольку локализуемая гравитация внутренне неполиномиальна, мы приходим к выводу, что влияние гравитации прежде всего проявляется через логарифм ньютоновской константы, а не в результате учета эффектов порядка $\alpha\varkappa_g^2$. Как это ни поразительно, но значение $\log \varkappa_g^2 m^2$ почти равно α^{-1} — факт, который мы пытались понять, постулируя, что (почти) вся собственная масса электрона, возможно, имеет электродинамическую (учитывающую наличие гравитации) природу.

3. С введением F-мезона и постулированием двутензорной теории гравитации ядерная физика оказалась бы другим наименованием сильной гравитации (при условии, что извне добавлены явления, связанные с понятиями изотопического и унитарного спинов). Старейшина современной физики Эйнштейн, который потратил последние годы своей жизни на поиски унификации всех сил природы, я уверен, будь он жив, почувствовал бы себя счастливым от этого достаточно непосредственного и изумительного объединения понятий физики наиболее сильных и наиболее слабых сил природы. Я считаю, что мы слишком долгое время отдавали гравитацию на откуп космологам. Я хотел бы надеяться, что на следующей Рочестерской конференции, кроме секций сильных, слабых и электромагнитных взаимодействий, у нас будет, как и на этой конференции, четвертая секция, посвященная гравитации и физике элементарных частиц. Я надеюсь также, что к тому времени нашим уважаемым председателем или каким-либо другим экспериментатором будет установлена квантовая природа гравитационного излучения, служащая нам необходимой предпосылкой на всем протяжении этого доклада.

ПРИЛОЖЕНИЕ

На основании свойств микропричинности Джаффе классифицирует теории поля как локализуемые и нелокализуемые. Приведем следующие примеры.

А. Локализуемые по Джаффе теории

$$L_{\text{вз}} = g : \phi^n \exp(\varkappa\phi) : \quad \text{(чисто неполиномиальный лагранжиан)}$$

или

$$L_{\text{вз}} = g : (\bar{\psi}\psi\chi) \exp(\varkappa\phi) : \quad \text{(смешанный лагранжиан).}$$

Точки « : » означают нормальное упорядочение.

Б. Нелокализуемые по Джаффе теории

$$L_{\text{вз}} = g : \phi^n (1 + \varkappa\phi)^{-w} :$$

или

$$L_{\text{вз}} = g : (\bar{\psi}\psi\chi)(1 + \varkappa\phi)^{-w} : \quad (w > 0).$$

Произведенное Джаффе разделение теорий основано главным образом на поведении при высоких энергиях двухточечной спектральной функции $\rho(p^2)$, которая спадает скорее, чем $\sqrt{|p^2|}$ в локализуемом случае, и медленнее в нелокализуемом случае. Джаффе показал, что *локализуемые теории микропричинны*; нелокализуемые теории не удовлетворяют этому условию. Кроме того, на основании строгой теоремы Глазера, Эпштейна и Мартина можно показать, что элементы S-матрицы (на массовой оболочке) в локализуемом случае в области высоких энергий должны подчиняться ограничениям, сформулированным Фруассаром. Гораздо меньше известно о нелокализуемых лагранжианах, не считая результата Штейнмана — Тейлора, утверждающего, что вопреки возможному нарушению причинности построение конструкции Лемана — Циммермана — Симанцика для S-матрицы может быть все же проведено, так же как и доказательства СРТ-теоремы и теоремы о связи спина и статистики.

Ефимов и, следуя ему, мы в наших работах в Триесте использовали по видимости нелокализуемый вариант киральной и гравитационной теории (параметризацию Вейнберга — Швингера в киральном случае и общепринятую параметризацию $g^{\mu\nu} = \eta^{\mu\nu} + \varkappa h^{\mu\nu}$ в случае гравитации). Но в наших работах не содержалось ничего, что запрещало бы нам использовать альтернативную экспоненциальную параметризацию полевых переменных в обоих этих случаях. Таким образом, мы остаемся формально в локализуемом классе теорий и гарантируем тем самым все имевшиеся преимущества — в частности, фруассарову ограниченность S-матричных элементов на массовой оболочке. Если теоремы эквивалентности справедливы, то представляется вероятным, что по видимости *нелокализуемый* вариант киральных теорий или гравитации на самом деле окажется *локализуемым* благодаря тому, что имеют место многочисленные сокращения.

ЛИТЕРАТУРА

1. *Гельфанд И. М.*, *Шилов Г. Е.*, Обобщенные функции, т. I, М., 1958.
2. *Güttinger W.*, Fortschr. Phys., **14**, 483 (1966).
3. *Gustafson T.*, Arkiv. Mat. Astron. Fysik, **34A**, № 2 (1947).
4. *Bollini C. G.*, *Giambiagi J. J.*, Nuovo Cimento, **31**, 550 (1964).
5. *Speer E. R.*, Journ. Math. Phys., **9**, 1404 (1968).
6. *Mitter P. K.*, Oxford preprint 18/70, 1970.
7. *Constantinescu F.*, *Bloomer R.*, Univ. of Munich preprint (Nucl. Phys., 1971).
8. *Volkov M. K.*, Commun. Math. Phys., **7**, 289 (1968).
9. *Volkov M. K.*, Commun. Math. Phys., **15**, 69 (1969).
10. *Volkov M. K.*, Ann. Phys., N. Y., **49**, 202 (1968).
11. *Salam Abdus*, *Strathdee J.*, Phys. Rev., **D2**, 2869 (1970).
12. *Lehmann H.*, *Pohlmeyer K.*, DESY preprint 70/26.
13. *Karowski M.*, Commun. Math. Phys., **19**, 289 (1970).
14. *Salam Abdus*, в книге Proceedings of the VIIth Coral Gables Conference on Symmetry Principles at High Energies (ICTP Trieste preprint IC/70/7).
15. *Lee B. W.*, *Zumino B.*, Nucl. Phys., **B13**, 671 (1969).
16. *Филиппов А. Т.*, препринт, Дубна, 1968.
17. *Ефимов Г. В.*, ЖЭТФ, **17**, 1417 (1963).
18. *Efimov G. V.*, Phys. Letters, **4**, 314 (1963).
19. *Efimov G. V.*, Nuovo Cimento, **32**, 1046 (1964).
20. *Efimov G. V.*, Nucl. Phys., **74**, 657 (1965).
21. *Efimov G. V.*, ICTP Preprint Trieste, Non-local quantum theory of scalar field, Parts, I, II and III, August 1969.
22. *Efimov G. V.*, CERN report № Th. 1087, 1970.
23. *Fradkin E. S.*, Nucl. Phys., **49**, 624 (1963).
24. *Fradkin E. S.*, Nucl. Phys., **76**, 588 (1966).
25. *Delbourgo R.*, *Salam Abdus*, *Strathdee J.*, Phys. Rev., **187**, 1999 (1969).
26. *Hunt A. P.*, *Koller K.*, *Shafi Q.*, Imperial College preprint ICTP/69/12.
27. *Fels S.*, Phys. Rev., **D1**, 2370 (1970).
28. *Fried H. M.*, Phys. Rev., **174**, 1725 (1968).
29. *Taylor J. G.*, Univ. of Southampton preprint, July 1970.
30. *Kapuścik E.*, ICTP, Trieste, Internal Report IC/69/132.
31. *Delbourgo R.*, *Koller K.*, *Salam Abdus*, Imperial College preprint ICTP/69/10 (Ann. Phys. N. Y., в печати).
32. *Keck B.*, *Taylor J. G.*, Univ. of Southampton preprint, April 1970.
33. *Okubo S.*, Progr. Theoret. Phys. (Kyoto), **11**, 80 (1954).
34. *Arnowitt R.*, *Deser S.*, Phys. Rev., **100**, 349 (1955).
35. *Volkov M. K.*, Commun. Math. Phys., **7**, 289 (1968).
36. *Shafi Q.*, Imperial College preprint ICTP/69/22 (Phys. Rev., в печати).
37. *Epstein H.*, *Glaser V.*, *Martin A.*, Commun. Math. Phys., **13**, 257 (1969).
38. *Sakita B.*, *Virasoro M. A.*, Phys. Rev. Letters, **24**, 1146 (1970).
39. *Delbourgo R.*, *Salam Abdus*, *Strathdee J.*, Letters al Nuovo Cimento, **2**, 354 (1969).
40. *Salam Abdus*, *Strathdee J.*, Lettere al Nuovo Cimento, **4**, 101 (1970).
41. *Isham C. J.*, *Salam Abdus*, *Strathdee J.*, ICTP Trieste preprint IC/70/131 (Phys. Rev., в печати).
42. *Isham C. J.*, *Salam Abdus*, *Strathdee J.*, ICTP Trieste preprint IC/70/108 (Phys. Rev., в печати).
43. *Zumino B.*, CERN preprint and Brandeis Lecture Notes, June 1970.
44. *Perveen S.*, Journ. Phys. A. General Physics, **3**, 625 (1970).

СОДЕРЖАНИЕ

УВАЖАЕМЫЙ ЧИТАТЕЛЬ!

Ваши замечания о содержании книги, ее оформлении, качестве перевода и другие просим присылать по адресу: 129820, Москва, И-110, ГСП, 1-й Рижский пер., 2 Изд-во «Мир»

КВАНТОВАЯ ГРАВИТАЦИЯ И ТОПОЛОГИЯ

вып. 2

Редактор Н. Л. Телеснин
Художник Г. А. Щетинин
Художественный редактор Е. Урусов
Технический редактор Е. С. Потапенкова
Корректор В. С. Соколов

Сдано в набор 26/II 1973 г.
Подписано к печати 6/VII 1973 г.
Бумага тип. № 1 $60\times84^{1}/_{16}$=6,75 бум. л.
12,56 усл. печ. л.
Уч.-изд. л. 13,15 Изд. № 2/6713
Цена 1 р. 32 к. Зак. 546

ИЗДАТЕЛЬСТВО „МИР"
Москва, 1-й Рижский пер., 2

Ордена Трудового Красного Знамени
Ленинградская типография № 2
имени Евгении Соколовой Союзполиграфпрома
при Государственном комитете Совета
Министров СССР по делам издательств,
полиграфии и книжной торговли
г. Ленинград, Л-52, Измайловский проспект, 29